

EXHIBIT 37

[Filed Under Seal]

Evaluation of the Taft Demonstration Project:
Performance of a Private-Sector Prison and the BOP



Federal Bureau of Prisons
320 First Street, NW
Washington, D.C. 20534

Harley G. Lappin
Director, Federal Bureau of Prisons

Thomas R. Kane, Ph.D.
Assistant Director, Information, Policy and Public Affairs

William G. Saylor
Director, Office of Research and Evaluation

Scott D. Camp, Ph.D.
Project Director

Contributors
Scott D. Camp, Ph.D.
Dawn M. Daggett, M.S.

October 7, 2005

CONTENTS

ACKNOWLEDGMENTS	v
EXECUTIVE SUMMARY	vii
SECTION I. INTRODUCTION	1
Outline of the Current Report	3
SECTION II. CONTRACTING FOR TAFT	5
SECTION III. RESEARCH LITERATURE ON PRISON PRIVATIZATION	7
Comparative Costs of Public and Private Prisons	8
Comparing Operations of Public and Private Prisons	12
Summary	34
SECTION IV. PERFORMANCE OF TAFT	35
Results from Models for One Time Point	39
Multilevel Models of Institutional Performance	45
Aggregate-Only Models of Institutional Performance	50
Note on Interpretation of Residuals	56
Random Drug Test Results	57
Misconduct at the TCI Minimum-Security Camp	58
Summary	59
SECTION V. CRITICAL INCIDENTS	77
Chronology of Incidents at TCI	78
Chronology of Events at BOP Comparison Prisons	82
FCI Forrest City	82
FCI Elkton	85
FCI Yazoo City	87
Summary	88
SECTION VI. CONCLUDING REMARKS	89
APPENDIX 1. MISCONDUCT CODES AND CATEGORIZATIONS	93
APPENDIX 2. ADDITIONAL MULTILEVEL PERFORMANCE MEASURES	99
APPENDIX 3. PERFORMANCE MEASURES FOR TAFT CAMP	107
REFERENCES	113

ACKNOWLEDGMENTS

This report was written with the support and assistance of many individuals at the Federal Bureau of Prisons. While it is not possible to name all of those who contributed, it is necessary to mention those whose assistance was invaluable. First, Kathleen Hawk Sawyer, former Director of the Federal Bureau of Prisons and Harley G. Lappin, current Director of the Bureau of Prisons, provided essential support for the project. Other key administrative support was provided by Dr. Thomas R. Kane, Assistant Director of the Information, Policy, and Public Affairs Division and Judi Simon Garrett, Deputy Assistant Director of the Information, Policy, and Public Affairs Division. Judi Simon Garrett can be contacted for additional information.

Key cost data and assistance were provided by Carol Durkee, Chief, Office of Budget Execution, and her staff. Patricia Bowersox, Management Analyst in the Correctional Programs Division, provided the data from the BOP 583 forms which covered serious incidents of inmate misconduct. Roxie Schoppet, Management and Program Analyst in the Health Services Division, provided data on the Public Health Service staff working at BOP prisons.

Key support was provided by members of the Office of Research and Evaluation. William (Bo) Saylor, Chief, Office of Research and Evaluation, took over the major oversight responsibility of the project upon the retirement of Gerald (Gerry) Gaes, former Chief of the Office of Research and Evaluation. Gerry also helped develop the theoretical understanding of privatization in the office, and Bo and Gerry collaborated on the methodological design of the evaluation. Other key support came from Francisco (Chico) Barroso, formerly a Computer Specialist in the Office of Research and Evaluation. Chico generated the data for the individual-level analyses of misconduct. Jennifer Batchelder, Social Science Research Analyst, helped put together the variables and data used in the aggregate-only models of inmate misconduct. Others who helped with programming tasks include Suzanne (Suzy) Vanyur, Stephanie Rogers, Neal Langan, Nancy Miller, and Sean Rosenmerkel. Susan (Sue) Allison and Victoria Joseph helped with the analysis of the 583 data.

The support of all of these individuals was key to the success of this project.

EXECUTIVE SUMMARY

In 1996, the United States Congress directed the Federal Bureau of Prisons (BOP) to operate the federal prison at Taft, California as a demonstration of prison privatization (see Conference Report to Accompany H.R. 3610, Making Omnibus Consolidated Appropriations for Fiscal Year 1997, Public Law 104-208). This report describes selected measures of prison performance for the Taft Correctional Institution (TCI), operated by Wackenhut Corrections Corporation, especially in comparison to three BOP prisons built upon the same architectural design at about the same time and holding similar types of inmates.¹

Contract Issues

The contract at TCI was designed on the ideas of performance-based contracting. To the greatest extent possible, the solicitation for work specified performance goals for the contractor, but the contractor was given the discretion to develop operational procedures, practices and policies that would achieve the objectives. While not a new idea in other areas of government, the performance-based contract at TCI was the first of its kind in corrections. The resulting contract differs from typical contracts for correctional services in other important ways. For example, the contract contains provisions for performance awards for performance that exceeds the requirements of the contract. This gives the contractor an incentive to go beyond the work specified in the details of the contract to implement sound correctional practices. Most contracts only contain provisions for withholding payment for noncompliance with the contract, and the Taft contract does contain these provisions as well.

The contract for TCI was designed to give the contractor flexibility in implementing operational procedures and an avenue for reward when implementing practices not specifically called for in the language of the contract. According to the privatization

¹Wackenhut Corrections Corporation or WCC changed its corporate name to The GEO Group, Inc. (GEO) in 2003 after it became a fully-independent company. Formerly, WCC was a subsidiary of Wackenhut Corporation, a private-security firm. The corporation is referred to in this report as Wackenhut, Wackenhut Corrections Corporation, or WCC.

administrator at the BOP, the structure of the contract is working as Wackenhut Corrections Corporation has received fewer deductions for services not performed over the course of the contract and greater amounts of the potential award fee (5 percent of the contract amount).

Prior Research on Quality of Private Prisons

Researchers in the Office of Research at the Federal Bureau of Prisons have published reports and a book demonstrating the proper methods for conducting evaluations of the quality of private prisons (Camp & Gaes, 2001; Gaes, Camp, Nelson, & Saylor, 2004; Gaes, Camp, & Saylor, 1998). The overriding theme of the writings is that most evaluations of private prisons have not separated management performance issues from other differences between prisons that affect the typical outcome measures examined, such as age of the prison, the mission of the facility, or the composition of the inmate population. In contrast, BOP researchers have shown how to separate performance factors from nuisance sources of variation in measures of performance derived from staff survey data, inmate survey data, and inmate misconduct data (Camp, 1999; Camp & Gaes, 2002; Camp, Gaes, Klein-Saffran, Daggett, & Saylor, 2002; Camp, Gaes, Langan, & Saylor, 2003; Camp, Gaes, & Saylor, 2002; Camp, Saylor, & Harer, 1997; Camp, Saylor, & Wright, 1999). Many of these reports included a focus on TCI and found that TCI generally performed at the lower-end of performance with respect to BOP low-security prisons.

Performance at the Taft Correctional Institution

Performance at TCI was primarily assessed by examining levels of misconduct among inmates from 1998 through the first half of 2003. Since misconduct data are influenced by policy decisions and the ability to detect misconduct in addition to the actual underlying level of misconduct, the results of random drug tests were also examined. Data on random drug tests provide an indication of the actual rate of drug use at the prison in question, and differences between prisons in positive hit rates are not confounded with factors that cannot be accounted for statistically.

Misconduct was examined by first looking at the overall level of all types of misconduct and then breaking down misconduct into categories of misconduct. There were two

categorizations of misconduct used, a substantive categorization of misconduct into type of misconduct (drug, violent, etc.) and a categorization based on the severity of the misconduct. The second categorization, based on severity, was the primary breakout used to assess performance at TCI.

The evidence suggested that there were performance issues at TCI. TCI had higher counts than expected for most forms of misconduct, including all types of misconduct considered together. The problems in misconduct were centered around the most serious of BOP misconduct (100-level offenses) and the least serious (300- and 400-level offenses). TCI performed more along the lines expected for 200-level offenses.

Consistent with the problematic trends for misconduct at TCI, the results of the analysis of positive results for random drug tests were also unfavorable at TCI. TCI had the largest deviation of observed from expected values for most of the time period examined. The deviation was positive, meaning that TCI had more instances of positive findings for random drug tests than would have been expected.

Critical Incidents at Taft Correctional Institution and BOP Comparison Prisons

The BOP takes all forms of inmate misconduct seriously. Some forms of misconduct require special attention because of their potential impact upon prison operations. Serious acts of misconduct, such as serious assaults or group events, are reported daily to BOP administrators in Central Office with a form known as a BOP 583. The most serious incidents at BOP prisons are examined with a process known as an after-action review.

A review of 583 forms and after-action reviews uncovered some instances of serious misconduct at TCI and the three BOP prisons that have been designated as comparison prisons because of similarities in mission, architecture, activation dates, and inmate populations. The most serious incidents occurred at TCI where there were two escapes from inside the secure perimeter fences—one in September of 1999 and another in December of 2003—and one general disturbance involving up to 1,000 inmates in November of 1999. During the same period of time, the BOP had three escapes from a secure prison, and the BOP operated in excess of 100 prisons during this time.

TCI had other incidents as did the BOP comparison prisons. The incidents at the BOP comparison prisons mostly involved fights involving 5 or more inmates which are classified by the BOP as group disturbances, although there were other actions such as work and food strikes. Many of the BOP group fighting incidents included gang activity, especially at FCI Forrest City.

Conclusions

The general issue addressed in this report pertains to the quality of operations at TCI. Quality of operations, as measured by contract performance, improved over the period examined here. At the beginning of the contract, contract deductions were more common, and the proportion of the potential award fees earned was lower. Over time, deductions decreased as award fees increased. This strongly suggests that the BOP was receiving the value that was expected from the operation of TCI by Wackenhut Corrections Corporation.

Relative performance at TCI, though, did not demonstrate the same improvement. TCI consistently demonstrated lower levels of performance on the performance measures examined here, primarily inmate misconduct and illegal drug use. This relationship holds both when TCI is compared to the three BOP comparison prisons as well as when TCI is compared to other BOP low-security prisons. TCI experienced three significant incidents that did not occur at the BOP comparison prisons. TCI experienced two escapes from inside of the secure perimeter of the low-security prison and one large-scale disturbance in which at least 1,000 inmates refused to return to their cells. These instances endangered both public safety and institution safety.

SECTION I. INTRODUCTION

In 1996, the United States Congress directed the Federal Bureau of Prisons (BOP) to operate the federal prison at Taft, California as a demonstration of prison privatization within the BOP (see Conference Report to Accompany H.R. 3610, Making Omnibus Consolidated Appropriations for Fiscal Year 1997, Public Law 104-208). BOP administrators took the language of “demonstration project” to include a formal evaluation of the success of the prison in comparison to similar BOP prisons after five years of operation. In one sense, there was nothing new in the directive from Congress. The BOP contracted with the non-profit and for-profit segments of the private-sector to hold federal inmates prior to the Taft directive. A prison holding illegal aliens at Eloy, AZ was operated by a for-profit firm in the private sector prior to Taft. More importantly, all prerelease, halfway houses holding federal inmates in transition back to their communities were operated by private sector firms prior to 1996. In addition, the BOP had three inter-government agreements (IGA’s) in place with local governments in Texas to hold illegal aliens. The local governments subcontracted the operations of these prisons to for-profit firms. The prisons are located at Big Spring, Eden, and Garza City. All of these arrangements with the private sector are still in place in 2005, and the BOP expanded the number of contracts with for-profit firms to include prisons in California City, California, Cibola County, New Mexico, McRae, Georgia, and Rivers, North Carolina. As of September 29, 2005, the BOP had 18,457 inmates in privately managed secure beds and 7,026 inmates in halfway houses operated by the private sector. Another private prison is being built in Philipsburg, Pennsylvania that will house an additional 1,500 male inmates.

Opening the Taft prison did break new ground in one important respect. The Taft prison, operated by Wackenhut Corrections Corporation (WCC) since 1998 and known as Taft Correctional Institution (TCI), remains the only secure, private prison to incarcerate general population inmates.¹ All of the other private prisons used by the BOP (excluding the halfway houses) incarcerate special populations of inmates, such as illegal aliens or

¹Wackenhut Corrections Corporation or WCC changed its corporate name to The GEO Group, Inc. (GEO) in 2003 after it became a fully-independent company. Formerly, WCC was a subsidiary of Wackenhut Corporation, a private-security firm. The corporation is referred to in this report as Wackenhut, Wackenhut Corrections Corporation, or WCC.

inmates sentenced by the District of Columbia. As such, TCI remains the only prison that resembles other prisons operated by the BOP that hold low-security inmates (the type of inmates held at TCI).

Healthy competition between the public and private sectors is often cited as one of the key benefits of privatization (Harding, 2001). The question before us is how that competition was translated into prison operations. In particular, it is necessary to examine the impacts of the introduction of prison privatization upon both prison costs and the quality of operations.² The BOP is an ideal natural laboratory for addressing these issues. First, the BOP is large enough that there are sufficient prisons to provide the context for developing the types of performance measures used here. Second, the BOP opened three other facilities during the period that TCI was opened using the same architectural design and holding similar types of inmates. Those prisons were located at Elkton, Ohio, Forrest City, Arkansas, and Yazoo City, Mississippi.

The BOP interest in obtaining information about the relative costs and quality of operations at TCI and comparable BOP prisons has been and continues to be multifaceted. The first prong of the strategy deals with the contract developed to guide the operations at TCI and the subsequent contract monitoring. As discussed in more detail in Section II, the BOP pursued a unique contracting strategy known as performance-based contracting, or contracting for results. The second prong of the strategy, put into place simultaneously with the contract award, was for the Office of Research at the BOP to devise and conduct research to compare operations and costs at TCI and comparable BOP prisons. With a decade of experience in performance measurement regarding quality of operations, the office immediately embarked upon research that would prove to be innovative in raising the bar for performance measurement in prisons. Lacking internal expertise, the office turned to a consultant to aid with cost comparisons. Dr. Julianne Nelson of The CNA Corporation was contracted to produce several reports for the BOP. The final prong of the research strategy involved contracting with objective, third-party researchers for an independent evaluation of the first five years of operations at TCI. This was accomplished by having the BOP transfer

²Prison costs are examined in a separate, companion report produced by Julianne Nelson (2005) of The CNA Corporation, *Competition in Corrections: Comparing Public and Private Sector Operations*. The report was commissioned by the Bureau of Prisons.

monies to the National Institute of Justice (NIJ). NIJ then awarded an initial contract for \$650,000 to Abt Associates, Inc. of Cambridge, Massachusetts following NIJ's normal competitive process.

Outline of the Current Report

The current analysis examines the quality of operations at TCI in comparison to similar BOP prisons. In so doing, this report covers analyses of TCI that have been previously published as well as research produced exclusively for this report. The new material on quality of operations primarily focuses upon institutional safety, in particular the ability to manage inmate misconduct. A more complete evaluation of prison operations would also include areas such as the quality of prison programs, covering topics such as education programs, health care provision (including mental health), recreation opportunities, and pro-social skills development.

The study progresses in the following manner. First, a general description of TCI is provided in Section II that focuses upon the contract that governs operations and contract monitoring. Next, a general review of the research literature and theoretical issues are presented in Section III. Previously published studies by BOP researchers are emphasized as they point to general limitations with other existing research on private prisons and provide examples of how different data sources can be used to construct performance measures to compare prisons. Section IV presents new information on using performance measures derived from information on inmate misconduct to view institutional performance over time. A description of significant incidents at TCI and the BOP comparison prisons is presented in Section V. This discussion provides a context against which to judge the more abstract discussions in other parts of the report. A summary of lessons learned from the evaluation of TCI is presented in Section VI.

SECTION II. CONTRACTING FOR TAFT

There are two general ways to structure a contract for operating a prison (Crane, 2001). First, the contract can be structured so that the means of operating the prison are incorporated into the contract language. Crane calls this measuring operations. In essence, the contractor is requested to operate the prison according to the policy and procedure controlling the operation of public prisons in the contracting jurisdiction. Second, the contract can be organized so that the contractor is encouraged to meet performance standards established by the contracting agency. Crane labels this approach measuring performance. As much as possible, the details about how to operate the prison are left to the contractor, although there are always importance exceptions to this general tendency.

The BOP adopted the latter approach, for the most part, in structuring the contract for the Taft facility. The BOP did specify that the contractor for the Taft prison had to meet the requirements established by the American Correctional Association (ACA) accreditation standards. The BOP uses ACA accreditation at its own prisons to establish a baseline of acceptable performance. In addition, the BOP imposed requirements on the contractor that key record keeping and tracking tasks – such as the inmate disciplinary process, case management, sentence computation, and inmate central files – be performed according to BOP program statements using BOP-developed software. These latter requirements were imposed to ease the record keeping tasks associated with moving inmates throughout the BOP system whether the receiving prison is public or private. Other than these requirements, the contractor was given the freedom to organize the operations at the Taft prison and to develop their own quality control plan that would then be monitored with a BOP-developed quality assurance plan (see BOP Solicitation No. RFP PCC-0001, 17 January 1997).

There were other features of the federal Taft contract that deviated from typical contracts awarded by state correctional systems. First, the Taft contract specified a management fee (a firm-fixed price service contract) that would be paid for the monthly operation of the Taft prison whether the institution held 1 inmate or 1,946 inmates (95 percent of designated bed space). Once the limit of 1,946 inmates is reached, the contract calls for extra payments (called the incremental unit price) to the contractor to cover marginal

costs. Most states pay contractors on a daily per diem basis. But basing payment on a per diem basis creates pressures to keep inmates at the private prisons even when keeping them there violates sound correctional practice. One head of corrections told the story of a disturbance at a private prison. When the state stepped in to expedite the movement of involved inmates to other facilities, the warden asked that the state keep the inmates in place. Typical correctional practice is to disperse inmates to other locations to separate them. The structure of the federal contract for Taft avoided such circumstances because there is no monetary incentive to hold inmates at the Taft prison.

The other key difference with the Taft contract was the inclusion of incentive payments (what the contract calls an award-fee) for “exceptional” performance above and beyond that specified in the contract. This provision provides incentive, albeit one fraught with risk on the part of the contractor, for the contractor to provide service levels that exceed the requirements of the contract. Again, it does not appear that any of the state contracts for private prisons provide such incentives for performance. More common in all contracts, including the Taft contract, is the practice of taking deductions for performance that does not fulfill the requirements of the contract.

Initially, the BOP placed six staff members at TCI to monitor the operations of the facility. While the BOP had long experience in monitoring inmates placed in secure private beds, these contracts were for a special segment of the BOP population—illegal aliens. The Taft facility was the first private prison to hold inmates who were not part of some specially defined population. For example, the BOP currently places many individuals sentenced under District of Columbia laws into private prisons to fulfill the language of the legislation placing these inmates under the custody of the Attorney General. Many monitoring functions once performed at Taft have been centralized as the BOP added more contract prisons.

The awarding of performance fees and the taking of deductions is contentious. The general trend has been toward greater contract compliance by Wackenhut Corrections Corporation (WCC). As reported by Michael Janus, former Privatization Administrator for the BOP, deductions declined from the inception of the contract until the current period. At the same time, the trend has been for WCC to receive a greater percentage of the award fee which has a maximum value of 5 percent of the yearly contract amount.

SECTION III. RESEARCH LITERATURE ON PRISON PRIVATIZATION

Proponents of prison privatization have long argued that private prisons can both operate more cheaply and with higher quality than public prisons (Crants, 1991). In more recent times, the argument has been modified to include the claim that prison privatization brings competition to the public sector and improves operations there (Harding, 1997). Opponents of prison privatization, on the other hand, usually put forward some variant of the argument that incarcerating citizens is ultimately a government responsibility that should not be turned into a profit-making venture. The courts have never ruled against the constitutionality of private prisons, so the arguments about government responsibility become more of a political or ethical appeal and need not concern us. More pertinent are the claims that private prisons do not necessarily save money nor do they provide quality of service.

In the discussion that follows, a brief overview of the avoided cost methodology used by Dr. Julianne Nelson, an economist under contract with the BOP, is presented. Dr. Nelson (2005), now at The CNA Corporation, produced a companion report, *Competition in Corrections: Comparing Public and Private Sector Operations*, that covers costs in great detail. The present discussion does not try to summarize or analyze existing cost comparisons. This type of analysis has been presented elsewhere (General Accounting Office, 1996; McDonald et al., 1998; McDonald & Patten, 2003; Nelson, 1998; Perrone & Pratt, 2003; Segal & Moore, 2002), although there is still dispute about whether private prisons produce significant cost savings. No significant studies of the costs of private prisons have appeared since the publication of these reviews. Instead, the discussion focuses upon the types of advances pioneered by Dr. Nelson in comparing the costs of public and private prisons, using well-established principles. A discussion of Nelson's methods moves us beyond the faults of previous studies to an understanding of the proper type of method for performing future cost studies.

Following this brief discussion on comparing the costs of prisons, a synopsis of the work done by researchers in the Office of Research and Evaluation (ORE) at the BOP on comparing the performance of prisons, including public and private comparisons, is

presented. This synopsis is presented within the context of existing studies that have compared the performance of public and private prisons, but again the intent of the discussion is not to present a blow-by-blow account of the shortcomings and strengths of previous studies. Instead, attention is given to how the work done by ORE extends upon the pioneering efforts of early researchers in this area, researchers such as Charles Logan (1990; 1992). Logan's work on privatization, along with the work of Saylor (1988; 1989; 1992) on performance measurement, provided the jumping off point for much of the research conducted by ORE.

Comparative Costs of Public and Private Prisons

Cost savings and cost shifting were a primary motivation for the return of private prisons in the United States beginning in the 1980s. Faced with rising budget deficits and increasing inmate populations, many jurisdictions found the need to rely on innovative techniques to increase the size of correctional systems. Prison privatization was one answer that many jurisdictions adopted, for a number of reasons. First, prison privatization made it possible to build prisons without resorting to a public referendum on bond issues. Instead, the capital costs of building the prison were borne by the private vendor, and the costs of both building and operating the private prison were mixed together and paid from the operational funds of the public entity. Depending upon the contract terms, ownership of the prison converted to the public agency after a specified number of years. Second, prison privatization held out the promise of states receiving better or comparable prison operations for much less cost. Although the claims are no longer as large as those heard in the 1980s, cost savings of 5 to 15 percent are still generally promised (Segal & Moore, 2002). A study financed by the Association for Private Correctional and Treatment Organizations found that states with private prisons spent almost 9 percent less on corrections than states without private prisons (Blumstein & Cohen, 2003).³ Finally, a survey of prison administrators uncovered that the major reason they pursued private prison beds was to reduce overcrowding, not to save costs. The prison administrators also noted that private beds were attractive because they could

³An interesting fact about the APCTO study is that three states were dropped from the analyses for what seems an arbitrary reason, state expenditures did not grow. The study also did not demonstrate whether spending levels became lower with the introduction of private prisons or whether they were lower all along.

be acquired more quickly than having the prisons built in the public sector, and they responded that the private beds provided operational flexibility (McDonald et al., 1998: 16).

The pressure to use private prisons eased somewhat as prison populations in many states stabilized around 2001, although there was an upturn in the overall growth rate in numbers of prisoners to 2.6 percent in 2002 (Harrison & Beck, 2003). North Carolina removed inmates from the two private prisons operated by Corrections Corporation of America in 2000 due to concerns about contract compliance. North Carolina officials were concerned for some time about staff positions remaining unfilled at the two CCA prisons in addition to perceived problems with prison security and inmate programs (Nathan, 2000). Most recently, Governor Napolitano of Arizona rejected the idea of building a private women's prison and instead decided to expand an existing state facility (Hart, 2003).

As might be expected, the studies comparing the costs of public and private prisons have generated controversy. Some researchers have gone so far as to claim that the private sector has largely missed the opportunity to realize significant cost savings in the United States because of their inability to penetrate the markets in the Midwest and Northeast where labor costs are highest given the greater presence of unions (Austin & Coventry, 2003). Instead, private prisons are predominantly located in southern states and to a lesser extent western states where labor costs at public prisons are already low. These same researchers also make the point that given the small presence of private prisons in the United States, that even if private prisons save money on a case-by-case comparison, the relative impact upon any state or federal government budget is relatively minor (Austin & Coventry, 2003), a point made earlier by McDonald and his associates (McDonald et al., 1998).

There are other issues associated with cost comparisons. Various reviewers have noted that the results of cost comparisons are specific to the prisons examined and do not generalize to prison privatization in general (General Accounting Office, 1996; McDonald et al., 1998; Perrone & Pratt, 2003). In addition, there is controversy over the quality of the comparisons (Thomas et al., 1996), especially the handling of certain costs in the comparisons, most particularly overhead costs (Nelson, 1998). The most general problem is that overhead costs of the contracting agency are not typically separated into

those costs that should be charged against the private prison (as well as the public prisons being compared) and those that should not be charged against the private sector facility. Some analysts go as far as to suggest that no overhead costs of the public agency letting the contract should be charged against the private prison, but this position is not defensible. Clearly, there are overhead costs associated with letting a contract, overseeing the contract, and maintaining ultimate accountability for the inmates in private prisons. These costs and others must be counted against the private prison if there is to be an accurate accounting of how much the state spends to incarcerate inmates in private prisons.

There is an intuitive simplicity to the notion that a private vendor only be charged for the amount of contract costs minus any receipts received from the vendor either directly or indirectly (such as commissary profits returned to the state or state taxes collected) when comparing the costs of private and public prisons. However, such an approach grossly underestimates the true costs of contracting for the services. As mentioned above, public-sector staff wrote the contract, awarded the contract, as well as monitored the contract once the prison was operating. Additionally, policies had to be written to cover the operations of private systems within the correctional system. All of these actions, and related tasks, had costs associated with them, but the costs are not always properly recognized when comparing the costs of public and private prisons. The point has been made most forcefully by Nelson (1999). Nelson claims that the private sector prisons have to be assessed a fair share of the overhead costs of a contracting agency, but only for those costs that cannot be avoided by privatization.

Nelson (1999: 3) explains cost avoidance in this way with reference to the BOP. An avoidable cost can be conceptualized in either of two complementary ways. First, avoided costs are “the additional costs that the BOP itself *would have incurred* if it had operated Taft as a Federal facility” (emphasis in original). Or, second, avoided costs are “the additional costs that the BOP *has avoided* by ‘outsourcing’ the operation of the Taft facility” (emphasis in original).

This method of thinking about the costs of contracting for a private prison provides a method for examining overhead costs to determine which costs were avoided and which were not. For example, some vehicle purchases for prisons are funded from accounts in the centralized administrative offices of the BOP in Washington, D.C. In Fiscal Year

1998, the BOP spent \$4,197,684 on these purchases. However, all of these vehicles went to prisons operated by the BOP, so the BOP avoided spending additional monies for vehicles at TCI by contracting with Wackenhut. Therefore, the costs for vehicle purchases as part of the overhead costs associated with contracting for TCI were totally avoided, and the total amount for vehicle purchases was subtracted from overhead costs that were allocated to TCI. The overhead for vehicle purchases was only counted against BOP-operated prisons. As another example, TCI was responsible for paying for the background investigations of its employees. As a result, the Central Office funds used for this purpose at BOP prisons (\$12,347,281) were not included in the overhead costs attributed to TCI, but they were included in the overhead costs attributed to BOP-operated prisons.

While many Central Office costs were avoided by contracting with TCI, Nelson found that other Central Office costs in FY 1998 were not avoided. In fact, of the \$115,312,279 in overhead spending at the BOP in FY 1998, only \$65,895,852 of the total overhead costs were avoidable, leaving \$49,416,427, or 42.9 percent, as not being avoided. One example of an unavoided costs is the salary of the director of the BOP. The Director of the BOP was ultimately responsible for the well-being of the inmates at the TCI prison in FY 1998, just as she was accountable for inmates at BOP-operated prisons. As such, the salary and benefits associated with the Director's position were split evenly across TCI and the BOP-operated prisons. Likewise, such costs as the rent paid for office space to house the director (and other Central Office staff) were not avoided by contracting for TCI. These costs were also split evenly across all prisons, including TCI.

Many other analysts conducting cost comparisons have relied upon formulaic solutions, such as adding 12 percent of personnel costs to the estimate of government operations as suggested by OMB (Office of Management and Budget) Circular A-76 (McDonald et al., 2002). At least in the case of the operations of the Taft facility for the BOP, such an approach overestimates the difference in overhead costs and unfairly disadvantages the government in cost comparisons. OMB Circular A-76 is intended for use in prospective decisions about whether to purchase the service from a contractor or provide it in-house when actual costs are unknown. The advantage of the analysis presented by Nelson is that overhead costs did not have to be estimated, they were known. Since the costs were known, they could be assigned to avoidable and non-avoidable categories, and the actual difference between the overhead assigned to government operations and private sector

operations could be calculated. The difference was much less than the 12 percent figure used by OMB. Assuming BOP operation of Taft in fiscal year 1998, 3.65 percent of operating costs needed to be added to the cost of government operations in order to reflect the avoided overhead costs associated with turning Taft over to WCC for operation. Expressed as a percentage of personnel costs instead of total operating costs, the correct adjustment to personnel costs was 4.8 percent, not the 12 percent advocated by OMB.

While it is tempting to treat the 12 percent figure as if it were written in stone since it is incorporated into Circular A-76, this was not the intention of OMB. There is a provision in Circular A-76 to obtain approval for a different value than 12 percent if proper documentation is presented. Even though this option is rarely if ever exercised, there is a clear recognition that the 12 percent figure is not a one-size-fits-all solution. Again, the advantage of retrospective studies like that provided by Nelson is that the actual overhead costs are known.

The pages devoted to quality of operations outnumber the pages devoted to cost comparisons in this section, but the disparate treatment does not reflect the relative importance of either topic. Instead, the longer treatment given to evaluations of performance or operations is provided because it is the primary focus of this report. The proper method for conducting a cost evaluation comparing public and private sector costs has been provided by Dr. Nelson (2005) in the companion report to this study, *Competition in Corrections: Comparing Public and Private Sector Operations*.

Comparing Operations of Public and Private Prisons

Researchers have reviewed evaluative studies of prison quality at public and private prisons and arrived at different conclusions despite examining essentially the same studies.⁴ On the one hand, Adrian Moore of the Reason Public Policy Institute (RPPI) concluded that private prisons do as well as public prisons and usually better on quality dimensions such as public safety, protection of staff and inmates, compliance with

⁴The present discussion focuses upon evaluations of prison quality in the United States. Most private prisons beds are located in the U.S., and a discussion of experiences in the U.S. is most appropriate for the purposes of examining the Taft prison.

professional standards, quality of programs, and quality of health care and food (Moore, 1998). On the other hand, researchers at the General Accounting Office and Abt Associates Inc. reviewed essentially the same studies and concluded that there was insufficient scientific evidence to conclude whether private and public prisons differed in terms of their general operations (General Accounting Office, 1996; McDonald et al., 1998).

Both the GAO and Abt studies recognized that there were two limitations to existing research. First, despite the ample opportunities for evaluating the performance of private prisons, little research has been conducted. Even the studies completed to date do not provide much information about the general issue of how well private prisons operate because the conclusions are limited to the specific prisons examined. The prisons that were compared were not selected in such a way as to make them representative of national or even state-wide trends. In scientific jargon, the conclusions reached in the studies did not generalize to other situations. In essence, then, the major point made by both the GAO and Abt researchers was that the scientific evidence was simply not available. Moore, on the other hand, took the more pragmatic, if less rigorous position that most of the existing studies concluded that private prisons were doing better than the public prisons.

The second source of disagreement regarding the existing evidence on quality of operations at private prisons pertains to judgements about the scientific rigor of the studies conducted to date. The strongest review to date of methodological problems with the studies reviewed by GAO and used by RPPI was provided as an appendix to a 1998 Abt report. The appendix was written by researchers in the Office of Research and Evaluation at the Federal Bureau of Prisons (Gaes et al., 1998). The BOP researchers concluded that the methods used in existing studies did not allow for a determination of whether observed differences between prisons were due to differences in management practices, differences in the types of inmates held, or differences in other factors not related to management performance – such as the age of the prison or the architectural design (Gaes et al., 1998). While the evaluators in the original studies typically concluded that differences arose because of superior management practices by private-sector operators, the actual methods employed by the evaluators did not permit such strong conclusions.

In an evaluation of prison operations in Louisiana, for example, Archambeault and Deis (1996) compared three prisons built at the same time in Louisiana. One of the prisons was operated by the state of Louisiana and the other two were operated by private prison firms, one by Corrections Corporation of America and the other by Wackenhut Corrections Corporation. This allowed for a comparison of operations at three prisons with a similar architectural design. Usually, evaluations of private prisons compare newer private prisons to much older public prisons. However, the prisons did differ in significant ways that influence inmate behavior, including the racial composition of inmates at the respective prisons. In other words, even though the three prisons had demonstrable differences in factors unrelated to performance that would impact the measures examined, and even though the authors recognized that these differences existed, Archambeault and Deis did not adjust for these differences in their analyses. Instead, their research design treated the prisons as though they differed in only random ways when this was clearly not the case.

There were other problems in the Archambeault and Deis analysis, including the substitution of statistical for substantive significance. For example, there were five escapes from one of the prisons operated by a private vendor, three from the prison operated by the other private vendor, and none from the state facility. Archambeault and Deis converted these escapes into monthly rates, compared the rates with a statistical technique known as analysis of variance, found that with so few instances of escape that the differences between the prisons were not statistically significant. In fact, even though they note that the public prison did well in managing escapes, they claimed that “each of the three prisons are fully meeting the obligation of protecting the public” (Archambeault & Deis, 1996: 121). Substantively, though, the differences in escapes at these prisons are quite meaningful to prison administrators as well as the general public.

A final type of problem found in the Archambeault and Deis study (1996) that is typical of studies comparing public and private prisons is the inclusion of comparison prisons that are clearly different from the prisons of interest, again, without any adjustment. In their study, Archambeault and Deis included comparisons between the low-security prisons operated by Wackenhut, Corrections Corporation of America, and the state of Louisiana to other prisons operated by Louisiana that were much older and with different prison missions, including comparisons to Angola, one of the best known high-security penitentiaries in the nation.

While the Archambeault and Deis study was singled out here, the criticisms apply equally well to other major evaluations comparing the quality of operations at public and private prisons. For example, Charles Logan produced a comparison of a private female prison in New Mexico with the old female prison formerly operated by the state as well as a female prison operated by the Federal Bureau of Prisons (Logan, 1992). Where Archambeault and Deis relied upon operations data collected by the state, most notably inmate misconduct data, Logan relied upon survey data collected from staff and inmates. Logan made a very similar methodological mistake. Even though he documented that the characteristics of staff and inmates differed across prisons, he did not adjust for those differences when making comparisons between the prisons. The comparisons were based upon the number of staff/inmates providing favorable evaluations of over 200 items. Since research has consistently demonstrated the importance of individual characteristics when evaluating prison conditions (Britton, 1997; Saylor & Wright, 1992), Logan should have used statistical adjustments to create the comparisons.

The present review focuses upon the empirical work done by ORE to develop rigorous methods for comparing the operations of public and private prisons. When using operational data on inmates or survey data collected from either inmates or staff to generate performance measures of prison operations, there are three key methodological points to consider. First, does it make sense to summarize the data about individuals into an institutional indicator? Second, does the measure need to be adjusted to account for the characteristics of the units being summarized? And, finally, are the actions or attitudes being measured influenced by the clustering of individuals within the respective organizational units, in this case prisons?

Not all individual measures have a direct connection to organizational processes. Lincoln and Zeitz (1980), for example, use the example of job satisfaction in their now classic article on aggregating individual-level measures. Job satisfaction clearly measures a trait of individuals, how happy or satisfied they are with the work they do. However, Lincoln and Zeitz argue that it is not clear what it means to take the average level of job satisfaction as a measure of an organizational property. Organizations do not have job satisfaction, even if organizations may contribute to job satisfaction.

Taking the argument the next step, even if there is a theoretical or logical connection between individual measures and aggregations of those measures, it still remains an

empirical question as to whether the aggregated measures express anything about the organization. Technically speaking, the question is whether the organization generates variability in the measures collected from individuals. In layman terms, it is a question of whether the same individual would provide different evaluations of the measure in question if they were located in different organizational units. For example, with inmate misconduct, the question is whether the level of inmate misconduct is pushed up or down by being located at different prisons. As we discuss below, there are statistical procedures known collectively as multilevel models that provide assessment of the organizational properties of aggregated measures, but almost no systematic evaluations of private and public prisons have employed the techniques with the exception of the work done by ORE researchers.

The second point has to do with adjusting aggregate measures for exogenous, or outside, factors that influence the measure in question. Take the case of the relationship between age and inmate misconduct. One of the most readily accepted findings in penology is that younger inmates are more likely to engage in institutional misconduct than older inmates. Assume that we are going to compare two prisons on the percentage of inmates involved in misconduct over the past six months. Unless the institutions are perfectly matched on the ages of their respective inmate populations, the percentage of inmate misconduct as a measure of prison performance is confounded with differences between the prisons in the distribution of inmates at different ages, with the institution having more younger inmates subject to an unfair disadvantage. The goal is to develop a measure that indicates the likelihood that inmates similar to one another become involved in misconduct at different prisons. Again, multilevel models permit this type of adjustment for differences in individuals as well as organizational differences, such as differences in architectural design or the inmate to staff ratio.

The final point is that aggregate measures created from individual-level data must account for what statisticians call interdependence among observations. Observations (in our case attitudinal and behavioral data provided by staff and inmates) are interdependent when the measure on one individual is affected by another individual also being examined. Traditional statistical techniques, such as those used in existing evaluations of private prisons, assume that observations are independent of one another. For example, it is extremely unlikely that any two individuals selected for a national survey know one another. Therefore, the information provided by each individual is completely

independent of the attitudes of other respondents. In the case of staff attitudes about prison or inmate misconduct, it is extremely likely that the individuals providing data *do* know each other, and there is some chance that their responses or behaviors are affected by this knowledge. Again, the point is that a statistical adjustment is needed to deal with the interdependence of observations, and while multilevel models do this, these models have not been employed in existing evaluations with the exception of the studies published by ORE researchers.⁵

Multilevel models have been described very clearly in the technical literature (Goldstein, 1995; Kreft & De Leeuw, 1998; Raudenbush & Bryk, 2002). For our purposes, we can simply note that multilevel models, also known as hierarchical linear models or mixed models, allow us to simultaneously handle data on individuals and the prisons at which the study participants work or are incarcerated. As such, multilevel models can be employed to address the three issues associated with aggregated measures: are the measures affected by organizational differences, do the measures need to be adjusted for differences among individuals providing the data, and are the measures impacted by interdependence among observations? If the three conditions exist, then we can aggregate the variables with multilevel techniques and construct appropriate measures on which to compare prisons.

The first study conducted by the ORE examined whether multilevel models could be used with survey data from staff at the BOP to compare prisons in terms of staff evaluations (Camp et al., 1997). The study demonstrated that some individual-level measures could be aggregated to create theoretically meaningful measures with sound measurement properties. For example, a scale measuring individual-level evaluations of institutional operations had meaning as an organizational measure. The multilevel techniques demonstrated that evaluations of institutional operations significantly varied across prisons even after controlling for the individual-level characteristics of the staff providing the evaluations. On the other hand, a scale of satisfaction with supervision was not shown

⁵Interdependence of observations causes a loss in statistical power in determining the statistical significance of observed relationships. This increases the likelihood of rejecting the null hypothesis and concluding that an effect exists for a given variable when in reality there is no effect (Type I error). The actual number of respondents providing independent information is smaller than the sample of individuals participating in the study, usually by an unknown amount.

to be useful as an aggregate measure, even though the scale was a good measure of individuals' satisfaction with supervision. Practices at different prisons were not sufficiently uniform to systematically raise or lower staff evaluations of satisfaction with supervision.

A companion article that appeared afterward examined in greater detail the consequences of not adjusting measures of organizational performance with multilevel models (Camp et al., 1999). The article presented graphical means of examining the mismatch between adjusted and unadjusted measures, a method that could also be used to present the results of multilevel models for use by nontechnical audiences. Two scales were examined in the article, a scale of job satisfaction and one for institutional commitment. As expected by Lincoln and Zeitz (1980), job satisfaction did not display any variation at the organizational level. In other words, the practices at different prisons had no impact upon how workers evaluated their satisfaction with their jobs. On the other hand, institutional commitment did vary systematically across prisons.

In a multilevel model where there is systematic variation in the average value, as in the case with institutional commitment, the model is known as a random intercepts model. The intercept in a regression type of model measures the average or expected value when all other variables are held at a value of zero. In a random intercept model, where there is systematic variation associated with each organizational unit, each organizational unit has an observed intercept which can be higher, lower, or the same as the expected intercept. The differences between the expected intercept and the observed intercept for each organizational unit are estimated in multilevel models with a technique known as empirical Bayes (EB) estimation.⁶

The top panel of Figure III.1 presents a graph of the empirical Bayes estimates of differences among prisons in how they affect staff evaluation of commitment to the institution. Prisons on the far left of the graph produced lower evaluations of institutional commitment, where prisons on the far right improved staff evaluations of institutional

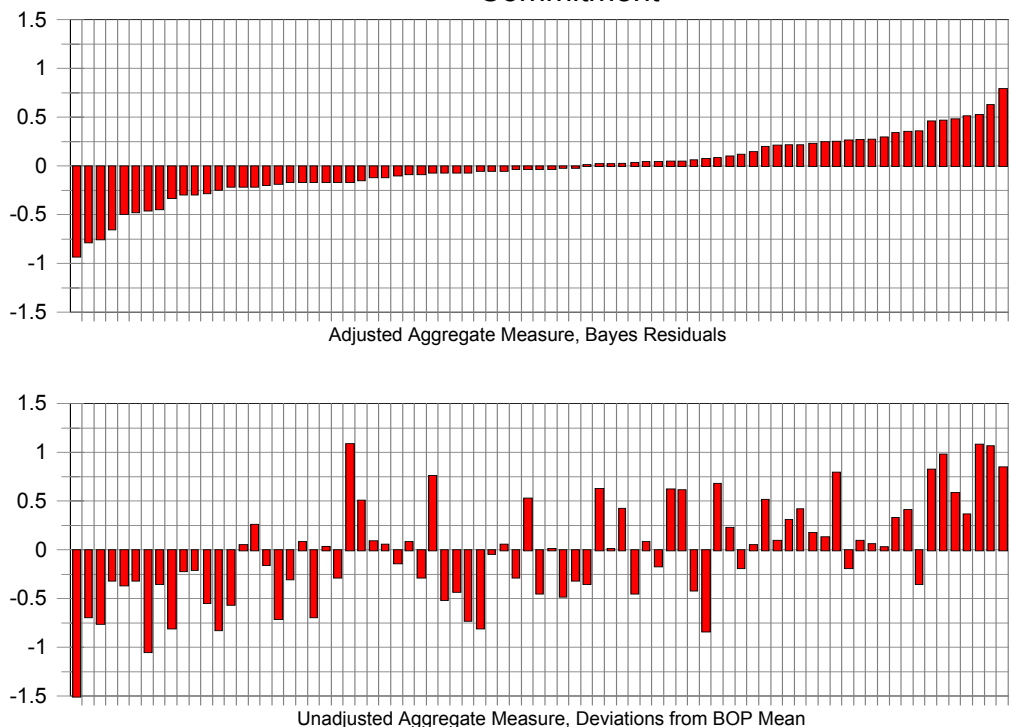
⁶There is another method used to estimate the residuals (the difference between the overall intercept and the intercepts for each organizational unit) implemented in other statistical software. This method is known as Markov Chain Monte Carlo estimation. The EB method is implemented in the software used for the analysis discussed here, the HLM software.

commitment. It is important to note that the graphs were produced from multilevel models that controlled for factors known to affect evaluations of institutional commitment (such as race, age, tenure, security level of the prison, etc.). The most a prison lowered staff evaluations of institutional commitment was about one point as seen in the graph. Since institutional commitment was measured on a seven point scale from low to high, this was a fairly large drop in average evaluations. On the other end of the scale, a couple of prisons raised evaluations of institutional commitment about .75 points, again on a 7 point scale. For confidentiality purposes, none of the prisons were identified by name in the published paper, but for operational purposes the graph would be produced with prison names attached.

Figure III.1 also gives an idea of the correspondence between organizational measures produced with multilevel models and organizational models that are not adjusted. In a manner similar to what has been used in existing evaluations of private prisons, the measures of organizational performance were computed as the raw averages of all respondents for the measure of institutional commitment.⁷ The specific prison average is subtracted from the average for all BOP prisons considered together to make the graph comparable to that presented for the empirical Bayes estimates. If there was no advantage to adjusting the measures, then the top and bottom panels of Figure III.1 would have been identical. Clearly, they are not identical, suggesting that relying upon unadjusted measures could produce fallacious conclusions.

⁷In some of the existing studies of private prisons, the evaluators computed the proportion or percentage of favorable responses. This is equivalent to what is done here, although our approach gives more precise information. Where we used the average of the scale score (from 0 to 6), these other studies used the average of favorable responses, where an unfavorable response received a value of 0 and a favorable response a value of 1. Scales such as institutional commitment were collapsed from 7 values to 2 values, 0 and 1. Taking the average of the 0's and 1's produced the proportion favorable in the other studies.

Figure III.1
Correspondence between Adjusted and Unadjusted Measures of Institutional Commitment



The measurement properties of the unadjusted organizational measures presented in the bottom panel of Figure III.1 were unknown because there are no methods for assessing them with this implied model. A complete discussion of the limitations of the implied model used by Archambeault and Deis (and others) can be found in the appendix to an unpublished report written by researchers in ORE (Camp, Saylor, & Gaes, 2001). The report is available upon request. The measurement properties of the measures generated from the empirical Bayes estimates presented in the top panel of Figure III.1 were examined. The variability of institutional commitment across prisons was significant and the ranking of institutions was reliable (Camp et al., 1999). Substantively, the evaluations of some prisons changed dramatically. For example, some of the prisons that produced slightly lower evaluations of institutional commitment on the adjusted scores had very

high scores of institutional commitment on the unadjusted measure. Had an evaluation of organizational performance been based upon the unadjusted measures, the evaluation would have concluded that one of the prisons had the highest level of performance observed (with average values of institutional commitment over a point higher than the BOP average). In reality, after adjusting for the characteristics of the staff and prison that produce favorable evaluations of institutional commitment, the prison actually produced lower evaluations than would have been expected.

The early studies investigating the use of multilevel models at the BOP were not concerned directly with comparing private and public prisons. However, the studies that followed did examine the performance of the Taft prison in comparison to BOP prisons. One study used staff survey data to compare operations at Taft and comparable BOP prisons (Camp, Gaes, & Saylor, 2002), another used inmate survey data to compare Taft (Camp, Gaes, Klein-Saffran et al., 2002), and a final study used BOP operational data to assess the effect of prisons upon inmate misconduct (Camp et al., 2003). The BOP researchers produced one other study that was based upon a subsample of federal inmates from the 1997 Survey of Inmates, administered by the Bureau of Justice Studies, but did not include information on the Taft prison (Camp, 1999). This study provided a no-cost assessment of whether it is worthwhile for the BOP to conduct their own inmate surveys to produce measures of organizational performance. Since the results confirmed that inmate responses to survey questions were influenced by prison differences, the BOP conducted its own survey of inmates in 1999 using the inmate version of the Prison Social Climate Survey developed by Saylor (1984).

The BOP administers the staff version of the Prison Social Climate Survey (PSCS) to its own prison employees every year. The survey started in 1988, and the response rate is usually around 90 percent as the BOP provides work time for employees to complete the survey. To date, the BOP has not administered the survey to workers at private prisons under contract with the BOP as part of its typical administration of the survey, but it did administer the survey to workers at Taft in 1998 and 1999. In 1999, the PSCS was also given to workers at a prison in Eloy, Arizona operated by Corrections Corporation of

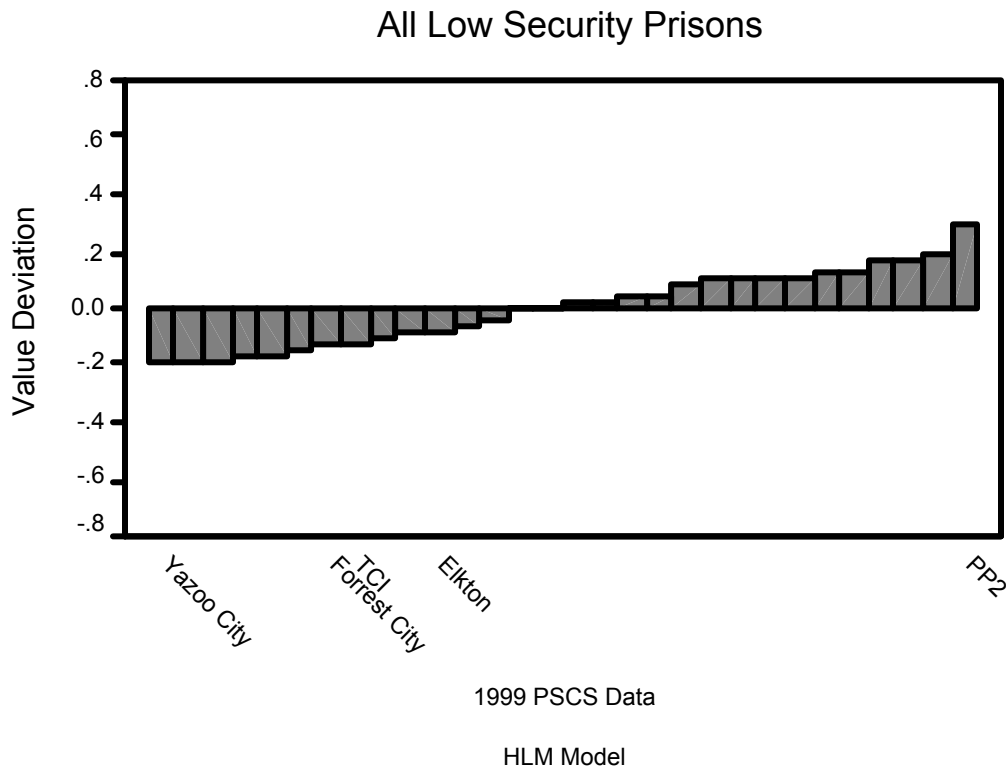
America for the BOP. Due to problems with administering the 1998 survey at Taft, only the 1999 survey provides data with face validity (Camp, Gaes, & Saylor, 2002).⁸

BOP researchers examined 17 different measures from the 1999 PSCS that covered areas such as assessment of management practices, the work environment, staff effectiveness, staff training, inmate safety, staff safety, and prison safety (Camp, Gaes, & Saylor, 2002). Only 4 of the measures had measurement properties that made them appropriate for creating aggregate measures to compare prisons, even though Logan had used many of these measures in his study (Logan, 1992). The appropriate measures were evaluations of institutional operations, organizational commitment, institutional commitment, and the perception of fire hazard in living units. Again, the study demonstrated that measures from individuals cannot always be aggregated to indicate differences between organizations.

The empirical results demonstrated that the performance at Taft was below the expected level on most measures, although Taft scored above the expected level on one measure. Figure III.2 provides information on institutional operations. Institutional operations is a scale derived from items asking respondents about items such as whether the institution is run well and whether the institution is viewed as being the best in the agency. All low-security prisons in the BOP as well as Taft are represented in the graph. The only prisons identified by name are Taft (TCI), FCI Elkton, FCI Forrest City, FCI Yazoo City. PP2 in the graph is another private prison, but it is not the focus of the present discussion. The three BOP prisons identified by name are the comparison prisons built at the same time as Taft on the same general architectural design. As can be seen in the graph, performance at Taft and the three BOP prisons tends to be in the bottom half of all low-security prisons in the BOP. Yazoo City actually produced the lowest evaluations of institutional operations of any low-security BOP prison. It is important to note that wardens and staff at the BOP comparison prisons have known all along that BOP

⁸In 1998, the Taft workers were supposed to receive a cover letter with their copies of the PSCS. The cover letter explained that the survey had numerous references to the BOP that the Wackenhut workers should treat as questions about Wackenhut Corrections Corporation. However, the cover letters were misplaced in the warehouse, and the workers received no instructions about how to treat questions about the BOP. In 1999, the survey was actually modified so that references about the BOP were changed to refer to Wackenhut Corrections Corporation at Taft and Corrections Corporation of America at Eloy.

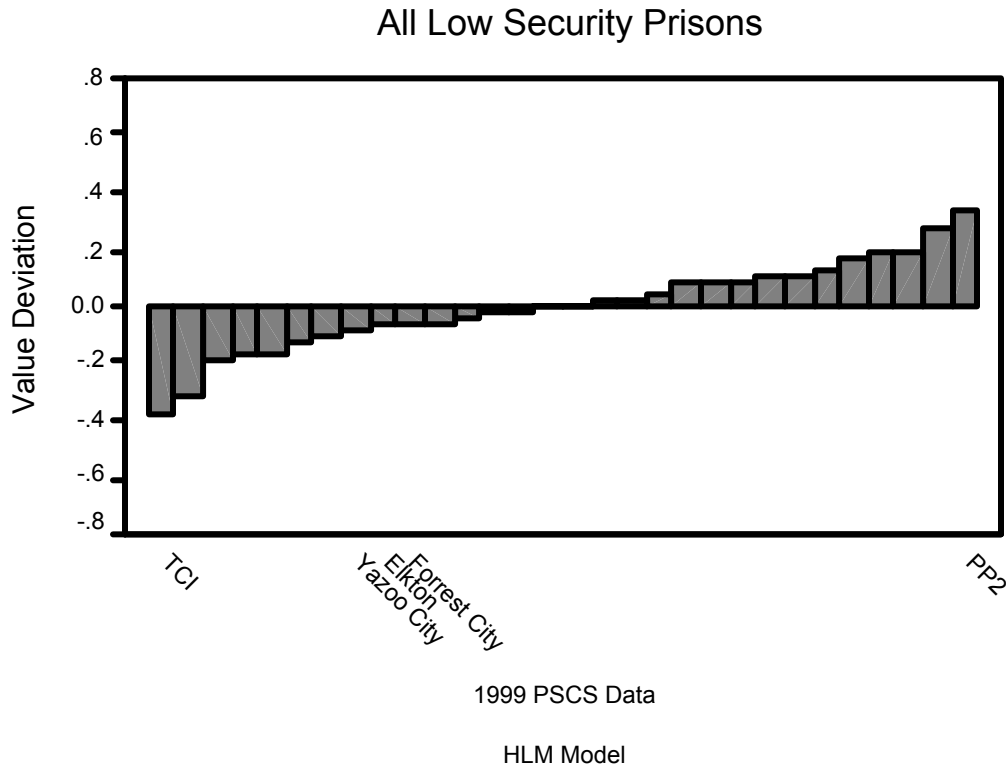
Figure III.2
Evaluations of Institutional Operations at Low-Security Prisons



management viewed them as comparison prisons for the Taft evaluation, although we have no systematic information on how this changed operations at these prisons.

Figure III.3 provides information on evaluations of organizational commitment. As can be seen there, Taft produced lower evaluations of organizational commitment on average than any of the BOP low-security prisons. While operations at Elkton, Forrest City, and Yazoo City did not push evaluations as low as observed for Taft, these prisons did have evaluations that were lower than expected given the characteristics of the prisons and staff working there. The scale of organizational commitment captures the degree to which an individuals identifies with the larger organizational, either the BOP or Wackenhut in this case, and intends to work for and remain part of the organization. Institutional commitment, discussed next, is a similar measure but in this case the referent is the

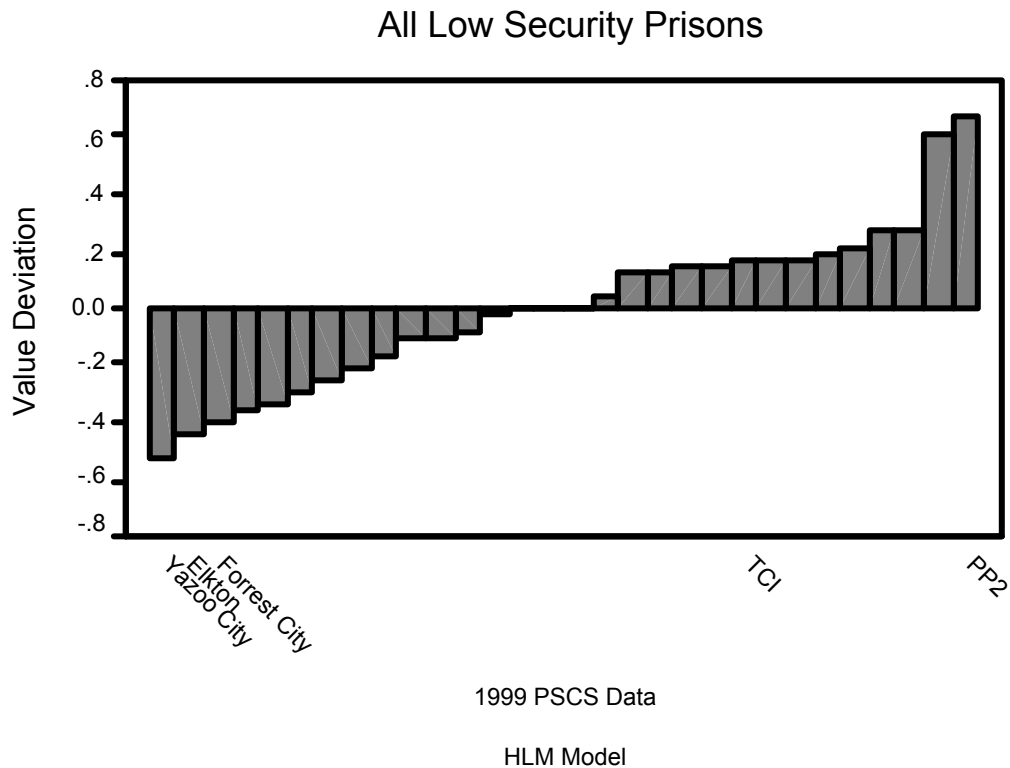
Figure III.3
Evaluations of Organizational Commitment at Low-Security Prisons



particular prison and not the larger organizational entity. If we focus exclusively on Taft and the three BOP comparison prisons, the BOP comparison prisons did “better” on this measure, although “better” still means pushing evaluations of organizational commitment to levels lower than expected.

Figure III.4 presents the results for evaluations of institutional operations. As is clearly evident in this graph, the three BOP comparison prisons did poorly on this measure as they produced the largest negative effects upon institutional commitment of any BOP low-security prisons. Taft, on the other hand, produced an effect that moved evaluations of institutional commitment higher than would have been expected, although not as high as some BOP low-security prisons and the other private prison indicated as PP2. On

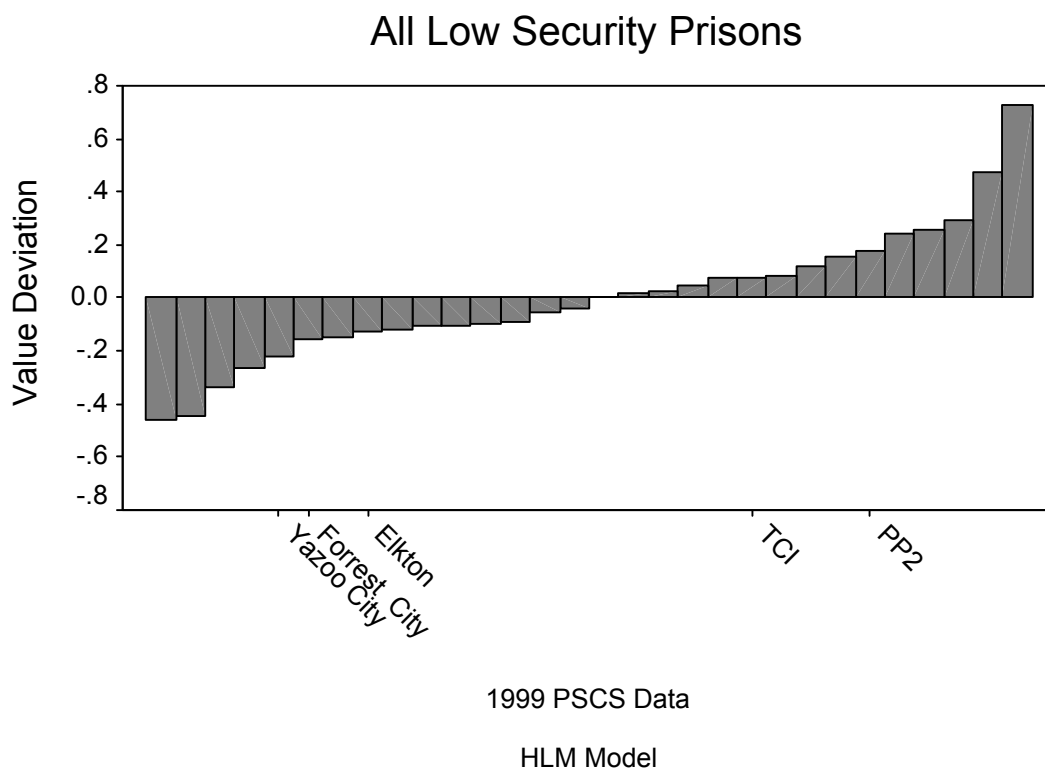
Figure III.4
Evaluations of Institutional Commitment



developing commitment to the institution among its employees, the operations or nature of the Taft prison certainly produced a favorable effect.

Figure III.5 provides information about how staff evaluated fire hazards in the living units at the respective prisons. This scale is reversed in direction from the previous scales discussed. That is, a higher score is an unfavorable finding because it indicates that staff believed there to be more of a fire hazard than would have been expected. It could be argued that the BOP comparison prisons did better than Taft on this measure as the perceptions of fire hazards were lower than expected at the comparison prisons and higher at Taft. While this is true, the evaluations of fire hazard in the living unit at Taft were not that different from what would have been expected.

Figure III.5
Evaluation of Fire Hazards in Living Units at Low-Security Prisons



As noted previously, the performance at Taft as measured by the indicators derived from staff survey data were mediocre at best, although Taft did generate favorable institutional commitment among its staff. Performance at the BOP comparison prisons was also lackluster in comparison to other BOP low-security prisons. It is not clear what produced these results, but it is worth noting that pay at Taft for line staff is higher than at other prisons operated by Wackenhut Corrections Corporation. The higher pay is a result of the federal contract which stipulates that the workers be paid according to amounts established by the Department of Labor. Regarding operations at BOP comparison prisons, there is little doubt that the different wardens viewed themselves as being in competition with Taft.

Following the 1999 study that confirmed that inmates were a reliable source to use in generating performance measures for prisons (Camp, 1999), the BOP administered the inmate version of the Prison Social Climate Survey in 1999 to ten male, low-security prisons. The prisons were not a random sample of low-security prisons, as Taft, Elkton, Forrest City, and Yazoo City were deliberately selected for inclusion because of the Taft evaluation. The other six prisons were selected to be representative of male, low-security prisons.

The results from the analyses of the inmate survey data surprised the BOP researchers in terms of the ability to use performance measures derived from inmate data. There was every expectation that the sample size of ten prisons would limit the ability to detect multilevel effects, especially the organizational-level effects, but this turned out to be much less problematic than expected. As much as possible, the inmate and staff versions of the PSCS use the same questions. Obviously, some questions asked of staff were not appropriate for inmates and vice versa. For example, it is not appropriate to ask staff about the quality of inmate food services or inmates about staff perceptions of staff safety.

One of the first steps of the analysis was to triangulate rankings of prisons provided by inmates and staff. Of the 15 measures chosen for analysis with an overlap between inmate and staff questions, only two measures were shown to have valid measurement properties for comparing staff and inmate evaluations (Camp, Gaes, Klein-Saffran et al., 2002). Totally unexpected, most of the measures were valid for inmate assessments but not staff assessments. The two questions where a comparison was possible were questions about sanitation (dirt) in the dining units and housing units. The organizational measures derived from inmates and staff correlated highly, .74 for the measure of sanitation in the dining units and .90 for sanitation in the housing units. A perfect correlation in this case would have produced a value of 1, so these correlations were quite high and also statistically significant. Despite concerns about the quality of information provided by inmates, it correlated quite well with information provided by staff.

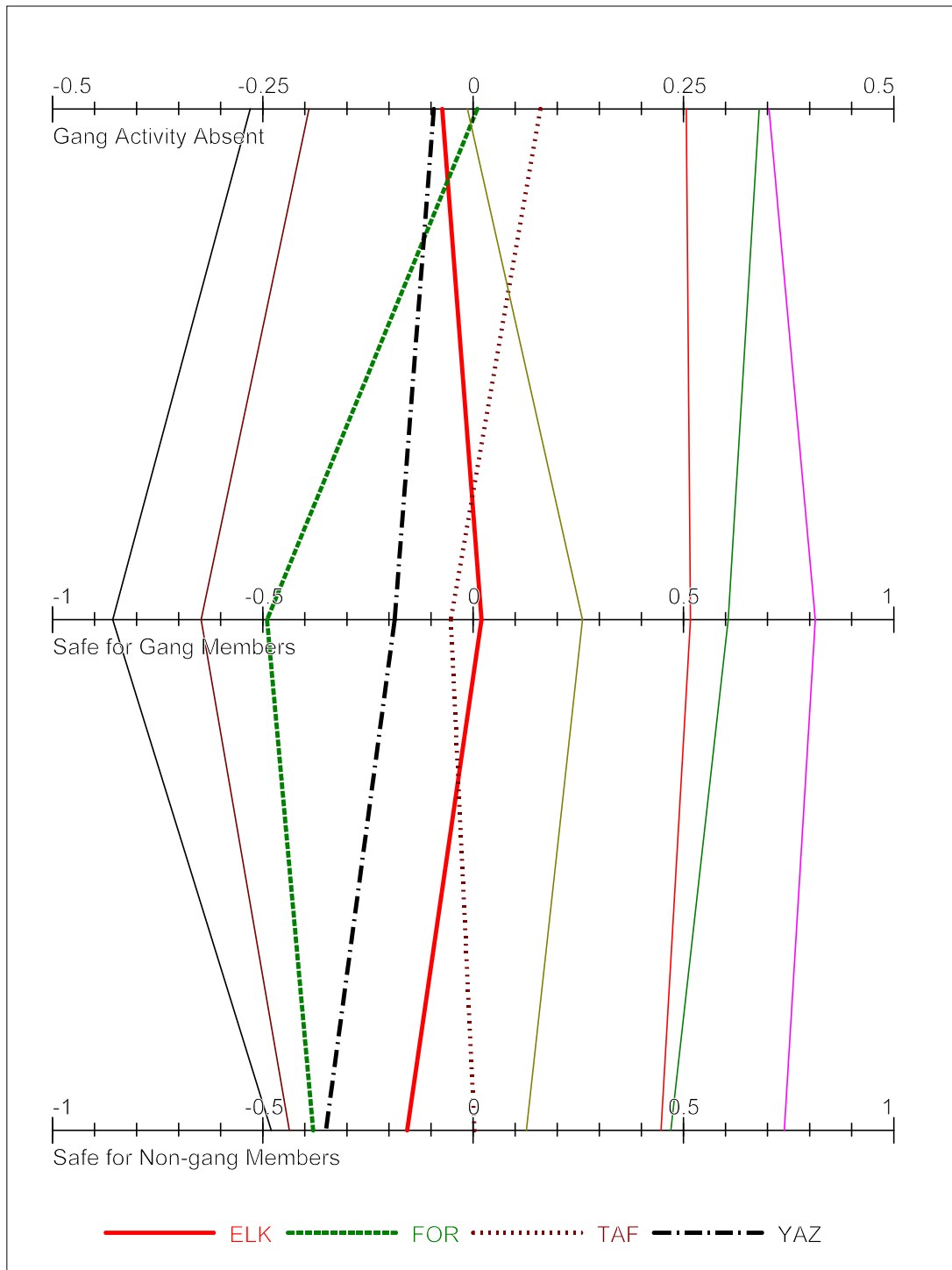
Part of the surprise from the inmate survey, then, was the fact that more measures could be used to compare prisons when the data came from inmates rather than staff. This is all the more surprising given the fact that there were only 10 prisons in the inmate survey and 96 in the staff survey. Generally speaking, more organizational units provide greater

reliability and validity for the organizational measures derived from the individual-level data. As such, the greater reliability of inmate scales suggested that inmate evaluations were more strongly influenced by the prison environment than were staff evaluations. Because of these strong measurement properties, the analysis was able to compare the 10 prisons on a host of measures using a technique known as parallel coordinates graphs.

Parallel coordinates graphs allow for multiple measures of related information to be presented together in one graphical form (Hartigan, 1975; Wegman, 1990). The software used to generate the graph presented in Figure III.6 was developed by Schoeneck Howell of the Office of Research at the Federal Bureau of Prisons (Howell, 2001). The information represents three measures of gang activity. Multilevel models were used to create adjusted organizational measures for each prison, the empirical Bayes residuals used in prior analyses. The values on the organizational measures for each prison were then plotted on the three axes, one for each measure of gang activity, and then the values for each prison were join with lines. In this graph, large values indicate positive performance, e.g., greater safety and absence of gang activity. The values for Taft, for example, are pretty close to 0 for all three measures indicating that inmates at Taft provided evaluations of gang activity that were at the expected level for BOP prisons. At the three comparison prisons, on the other hand, the results were not as favorable as inmate evaluations were lower than expected in terms of the prisons being safe for gang members, non-gang members, and in suppressing gang activity.

The paper also examined five related measures of security and safety, four measures of sanitation, and four measures of food services delivery. Three additional graphs where the measures were grouped according to substantive area appeared in the published paper (Camp, Gaes, Klein-Saffran et al., 2002). For sake of brevity, these additional graphs are not presented here. An examination of the graphs demonstrated that Taft had problems with respect to sanitation in the living units and dining halls in comparison to other BOP low-security prisons, and it had the lowest or next to lowest evaluations of food services. Evaluations of security, though, placed Taft in the middle of the pack. In sum, it did appear from the inmate evaluations that Taft had some issues, but the negative evaluations were not uniform across the substantive areas examined.

Figure III.6
Gang-Related Activity at Selected Low-Security Prisons

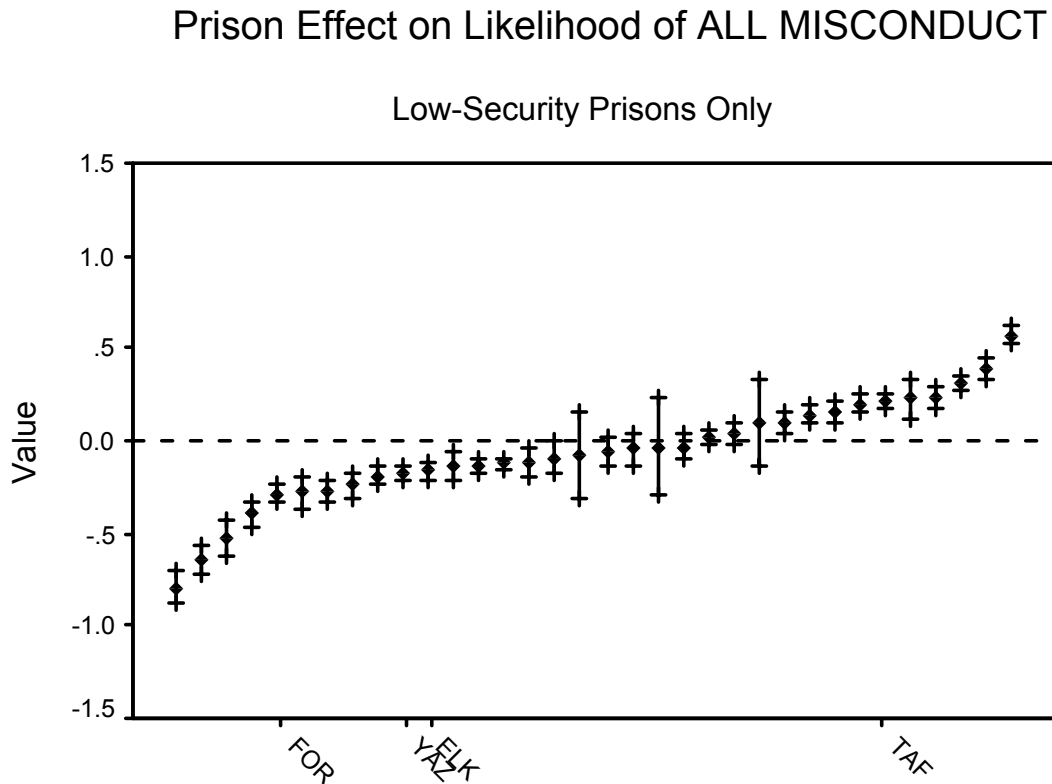


The BOP researchers also used the multilevel models to examine a cross-sectional view of inmate misconduct (Camp et al., 2003). This modeled approach stands in stark contrast to the unadjusted rates examined by Archambeault and Deis (1996) in their study in Louisiana. In the study, the BOP researchers demonstrated that it was necessary to control for differences across prisons in the socio-demographic characteristics and criminal histories of inmates incarcerated at different prisons, both at the individual and group levels. They also demonstrated that different categories of misconduct were not equally influenced by differences in practices across prisons. Sexual misconduct, for example, demonstrated no variability that could be associated with differences in prison practices.

When all types of misconduct were considered together, a fairly reliable organizational measure of prison performance resulted. Figure III.7 presents the results for all BOP low-security prisons. The graph in Figure III.7 is often referred to as a caterpillar plot, but it is the same information we have looked at in previous graphs. The empirical Bayes residuals from multilevel models measure whether a particular prison decrease or increase the probability that an inmate becomes involved in misconduct. Prisons whose values fall above the reference line of 0, such as Taft, had increased likelihood of inmate involvement in misconduct. Elkton, Forrest City, and Yazoo all produced a negative impact upon the likelihood of inmate misconduct.

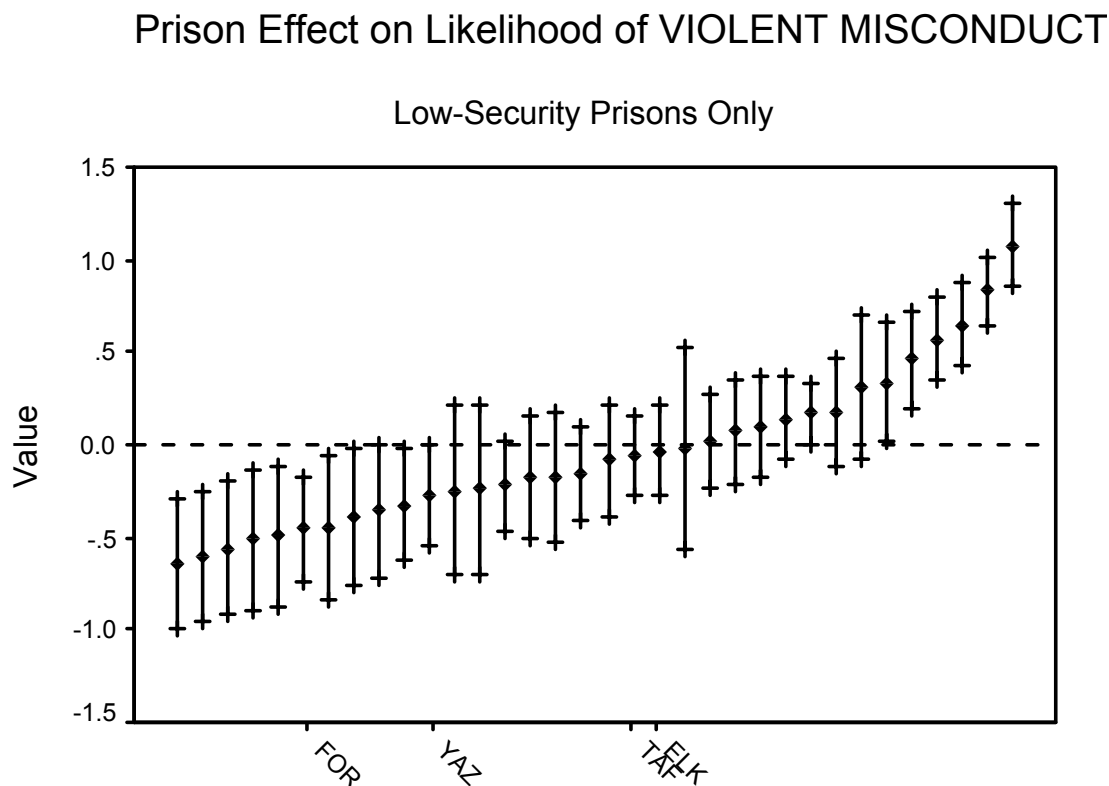
The new piece of information in the caterpillar plot of Figure III.7 is the confidence interval placed around each empirical Bayes estimate of organization performance. In a statistical model, all parameters are estimated with some degree of imprecision. The confidence interval gives the range in which it is believed the true value falls. In layman's terms, the importance of the confidence interval is that when comparing prisons, it is not sufficient to see if the midpoint for one prison is higher than another, it is also necessary to make sure that the confidence intervals do not overlap. If there is overlap in the confidence intervals, then it is not possible to claim that one prison is doing better/worse than the other.

Figure III.7



Since the measure derived from all forms of misconduct has high reliability, the confidence intervals presented in Figure III.7 are relatively small. On the other hand, a caterpillar plot for violent misconduct is presented in Figure III.8. Since violent misconduct is less common, it has a lower level of reliability as a measure of organizational performance. This shows up in Figure III.8 as larger confidence intervals around the empirical Bayes estimate of organizational performance. Unlike the case for all misconduct where Taft was more conducive to inmate misconduct than the BOP comparison prisons, in the graph of violent misconduct, we see that Taft and the three BOP comparison prisons all fall below the reference line of 0. However, only Forrest

Figure III.8



City can be said with any certainty to fall below the expected value for BOP prisons as the confidence interval does not include 0. The confidence intervals for Taft, Yazoo City, and Elkton all cross 0, indicating the possibility that practices at these prisons did not produce levels of violent misconduct different than would be expected otherwise.

In addition to the previously discussed investigations into methods for comparing prisons, researchers at the BOP conducted one study at the request of the U.S. Congress that is of relevance to this discussion (Camp & Gaes, 2000). A shortened version of the full report was later published in an academic/policy journal (Camp & Gaes, 2002). The BOP was asked to provide information on the growth of private corrections, standards for staff training in the private sector, and policy and procedure for custody practices among

private operators (Public Law 105-277, Sec. 111). The BOP contracted with Abt Associates to administer a survey to the contract administrators for the 103 contract prisons identified as operating in the U.S. or U.S. territories in 1999. Information was obtained from 91 of the 103 sites for a response rate of 88 percent.

Generally speaking, the survey documented that training standards are generally set by the contracting agency to be the same as those in the contracting jurisdiction. Regarding public safety issues, though, there were some problems noted in the private sector. In particular, the private prisons had much higher hit rates for illegal drug use, and the private operators had a much more serious problem with inmates escaping from secure prisons. From media reports, it was determined that private-sector prisons had 18 escapes in 1999 while holding slightly fewer than 70,000 inmates.⁹ The BOP held slightly more than 80,000 inmates in comparable prisons in 1999 and had one escape, the first in a number of years (Camp & Gaes, 2002).

The BOP researchers also documented that the workforce in federal prisons was much more stable than in private prisons. Between February and July of 1999, the BOP separation rate (for all reasons) was 4.4 percent. In contrast, 95 percent of the private prisons replaced staff at a rate of 10 percent or higher. In fact, over 20 percent of private prisons in operation for at least one year replaced 50 percent or more of their staff during this time period. While not examined by the BOP researchers, one problem caused by large amounts of turnover is having positions unfilled for lengthy periods of time. A recent audit of practices in the state of Tennessee uncovered this problem in the two prisons operated by Corrections Corporation of America for the Tennessee Department of Correction (Comptroller of the Treasury, 2003). In the period between November of 2000 and February of 2002, the South Central Correctional Facility operated by CCA had anywhere from 14 positions unfilled to 59 positions unfilled. With 59 positions unfilled, this meant that CCA lost the equivalent of 1,244 workdays for these positions. Even with 14 positions unfilled, 183 working days were lost (Comptroller of the Treasury, 2003: Appendix 5, page 76).

⁹The contract administrators actually reported 23 escapes for 1999.

Summary

Researchers at the BOP or under contract to the BOP have developed methods that allow for rigorous comparisons of public and private prisons on both cost and quality of operation. The quality of operations literature produced by ORE is among the first of its kind in the field of corrections, although the methods are used in other disciplines to great advantage (Normand, Glickman, & Gatsonis, 1997). Also, most of the ORE literature has been published in peer-reviewed journals. The cost-avoidance methodology employed by Nelson to examine the costs of operations is well established.

The evidence produced by the cost and quality studies for Taft suggest that the cost of operating Taft was comparable whether Wackenhut operated Taft or the BOP operated the prison. Complete information on the comparative cost of operating Taft is presented in the companion report to this report (Nelson 2005). The quality studies conducted to date by researchers at the BOP suggest that TCI is performing, for the most part, among the lower performing low-security prisons. TCI is not worse than all BOP low-security prisons, but it is also not among the best performers. The BOP comparison prisons also tend to be among the lower performing prisons on the measures examined, although they often compared favorably to TCI.

SECTION IV. PERFORMANCE OF TAFT

The mission statement of the Federal Bureau of Prisons notes: “It is the mission of the Federal Bureau of Prisons to protect society by confining inmates in the controlled environments of prisons and community-based facilities that are safe, humane, cost-efficient, and appropriately secure, and that provide work and other self-improvement opportunities to assist offenders in becoming law-abiding citizens.” It is fair to evaluate the performance of the BOP on these stated goals, a point that has been made elsewhere (Gaes et al., 2004). A complete appraisal of the points enumerated in the BOP mission statement would be far-reaching and would include evaluations of prison programs (work, educational, self-improvement), inmate health and safety, misconduct both within prison and after release, and other features of incarceration. This section focuses primarily upon one key aspect, inmate misconduct. More so than post-release behaviors, inmate conduct in prison is directly under the supervision and control of prison administrators and staff. Also, inmate misconduct speaks to the underlying reason for prisons, which is to confine inmates away from society to protect the public safety while simultaneously providing a safe and humane environment for inmates. While not examined here, the importance of prison programs for improving the odds for successful post-release integration in the community is ever more apparent with the current focus upon inmate re-entry. Unfortunately, research has lagged behind the need to evaluate these programs, and good intermediate measures of program success which can be collected in a timely fashion do not generally exist. Instead, research has focused upon the final outcome, e.g., the contributions of these types of programs to post-release recidivism (Gaes et al., 2004) or upon simple counts of inmates in programs. Recidivism data, though, must be collected over a number of years, and counts of inmates in programs without some assessment of program quality can be misleading.

Inmate misconduct data possess an undesirable characteristic from the viewpoint of being ideal for social-scientific measurement. Namely, misconduct data are susceptible to manipulation, and not necessarily manipulation that is part of an intent to deceive. Many types of misconduct in prison involve infractions of prison rules for the orderly running of the institution. Where one correctional worker may rely upon writing “tickets” (prison jargon for a formal disciplinary infraction) to insure prison order, another worker may be better at persuading inmates to comply with rules and rely less upon the formal

disciplinary process. Both workers may be equally effective in enforcing prison order, but the two workers would “generate” different amounts of official inmate misconduct. Of course, with the more serious types of misconduct, there is less discretion to handle the situation in a manner other than the officially proscribed one.

Another problem with inmate misconduct data is the extent to which the universe of all inmate misconduct is known. A prison which is poorly operated can have a low rate of inmate misconduct if the workers at the prison are unaware (intentionally or not) of most instances of inmate misconduct. Alternatively, workers at prisons can be encouraged to either focus upon or ignore certain types of inmate behavior that result in official misconduct, in much the same manner that police officers and prosecutors change emphasis and resources from one type of crime to another based upon changing directives. Even in a controlled prison environment, it is not possible for prison workers to be aware of every instance of inmate misconduct.

Given the potential issues with inmate misconduct data, we also examine the results of random drug tests at TCI and the BOP prisons. Unlike misconduct data, the collection of which are driven in part by the amount of the underlying behavior at a prison but also by administrative decisions and the ability to detect, random drug testing is specifically designed to measure the *prevalence* of the behavior at each prison. In addition to being designed to estimate the amount of the behavior at each prison, data on random drug tests are desirable because they speak to the basic functions of a prison. That is, the ability to prevent illegal drug use reflects the ability of the prison administration to implement security and custody procedures to prevent the infiltration of illicit drugs from the outside as well as the ability to observe and develop intelligence about the use and distribution of the illicit material once within the prison perimeter. As will be demonstrated, though, even random drug data are susceptible to manipulation if the procedures for testing and reporting are not followed.

In Section III of this report, the importance of developing performance measures that are adjusted for factors not related to performance was emphasized. The research presented in this section continues to build upon this idea. Performance measures are constructed from inmate misconduct data using two techniques. The first technique is the multilevel method used in many of the studies reviewed in Section III. As noted there, multilevel techniques allow for the simultaneous examination of individual-level and

organizational-level data, and in most cases this technique is needed to provide properly adjusted measures (Heinrich & Lynn, 2000). Specifically, the multilevel models are intercepts as outcomes models. That means that the intercept is the only variable that is allowed to vary randomly. The intercept is the part of the model associated with average differences between prisons. Since the outcome is the probability that an inmate was involved in misconduct during the month in question, the models were technically specified as follows with software from the HLM 4 program.

$$\text{Level 1:} \quad \log \left[\frac{\text{prob}(Y_{ij}=1)}{1-\text{prob}(Y_{ij}=1)} \right] = \beta_{0j} + \Sigma \beta X_{ij}$$

$$\text{Level 2:} \quad \beta_{0j} = \gamma_{00} + \Sigma \gamma W_j + u_{0j}$$

The specific variables used at the individual level (the X_{ij}) and at the aggregate level (the W_j) are discussed in a later section of this chapter, “Multilevel Models of Institutional Performance.” In reality, a model is estimated that simultaneously combined level-1 and level-2 variables, so that the combined equation takes the form that follows. As discussed in the multilevel models section of the chapter, the u_{0j} terms are the deviations up or down in inmate misconduct associated with the j institutions. Technically, the measures are empirical Bayes estimates of the average institution effect.

$$\text{Combined:} \quad \log \left[\frac{\text{prob}(Y_{ij}=1)}{1-\text{prob}(Y_{ij}=1)} \right] = \gamma_{00} + \Sigma \beta X_{ij} + \Sigma \gamma W_j + u_{0j}$$

The second technique examined is a regression approach that uses data collected only at the organizational level, e.g., not on specific inmates. Even though we recognize that multilevel models are preferred in theory when aggregating individual-level measures into organizational-level measures, including measures such as inmate misconduct, there are practical reasons to examine how well models based only on organizational-level data perform.

The first practical reason for considering aggregate-data models is the nature of the data archived by the BOP for over a decade in the Key Indicators/Strategic Support System

(KI/SSS). All of the data collected there, which are easily available for further analysis such as that pursued here, exist only at the aggregate level. While individual-level data for inmate misconduct exist in tape archives separate from the KI/SSS system, these data are not pre-cleaned and as readily accessible to researchers inside or outside of the BOP. Therefore, timeliness is an issue. The second practical consideration is that the analysis of the institutional performance of Taft Correctional Institute conducted by Abt Associates Inc. is based upon aggregate data. Even though the Abt analysis had not been seen or discussed before the initial draft of this section, it was known that the Abt researchers only requested the aggregate data contained in the KI/SSS system from the BOP. We wanted to produce an analysis that is based on the same limitations as the Abt analysis for comparative purposes. We also wanted to develop an understanding of the limitations of models based upon aggregate data in this instance. The analysis presented in this section accomplishes all of these goals. A third practical reason to examine the aggregate data has to do with coverage of the instances of misconduct. The multilevel analysis of misconduct was only conducted for sentenced inmates. The reason for this limitation was that specification of the multilevel models depended upon the criminal history of the inmates, and such information, captured in an initial custody classification of the inmates, only existed for sentenced inmates. On the other hand, the aggregate data analyzed with the negative binomial regression models include all inmates and all instances of misconduct.¹⁰

Since the data modeled at the aggregate level are count data, a negative binomial was the appropriate specification for the regression. While the specific variables included in the model are discussed in the section titled “Aggregate-Only Models of Institutional Performance,” the general form of the model was as follows.

$$\log(\mu) = \beta_0 + \sum \beta_j X_j + \epsilon$$

The equation says that the rate, μ , is a function of the independent variables, the intercept, and the overdispersion parameter ϵ . The overdispersion parameter accounts for the typical underestimate of dispersion in the count data when a Poisson distribution is

¹⁰The practical consideration concerning coverage is not that crucial because the multilevel data analyzed for the 2003 article covered almost 95 percent of the inmates included in aggregate-only models (Camp et al., 2003).

used. Whether the overdispersion parameter is significant is easily tested. The residual analyzed for each institution is the difference between the observed count and the expected count given the model. In fact, the standardized Pearson residuals were used as the indicators of performance for each institution. A count higher than expected signified rates that were problematic and counts lower than expected indicated favorable performance.

The analyses presented here allow for an assessment of whether count models of misconduct capture the same information as individual-level models. That is, are deviations from expected counts of misconduct at a prison related to factors operating on the likelihood that individual inmates are involved in misconduct. While this point seems to follow with common sense, there are sufficient instances of inappropriate conclusions drawn from aggregate data, especially regarding the ecological fallacy and Simpson's paradox, that testing this assumption is warranted.

The ecological fallacy occurs when data that are measured on groups are used to make conclusions about individual behavior. In this case, data on prisons are used to make conclusions about the misconduct of individual inmates. Data at the aggregate level can never answer this question directly. Simpson's paradox arises when relationships for specific groups are obscured when combined with other groups. For example, imagine that men had higher wages on average than women in a fictional company, even though women had higher salaries than males for every occupational category in the company. How was this seeming paradox explained? Men were more concentrated in higher paying occupations than women.

Results from Multilevel and Regression Models for One Time Point

Previous research conducted by the Office of Research and Evaluation (ORE) demonstrated the feasibility of using survey data collected from both staff and inmates as well as inmate misconduct data in conjunction with multilevel techniques to generate organizational measures with desirable and understood measurement properties. As noted above, creating performance measures from databases containing both individual- and aggregate-data is more problematic for practical reasons than creating the same measures from aggregate data. In the discussion that follows, we compare the implications for using the different methods for generating performance measures for one point in time,

June of 2001. Following the examination of the respective models for one point in time, we look at longitudinal analyses of performance measures created with the different techniques to evaluate the feasibility of aggregate-only models but more importantly to evaluate the performance of TCI and the BOP comparison prisons.

The basic strategy to compare the results from the multilevel and regression approaches is to see how well different methods of constructing performance measures correlate in terms of ranking the performance of the respective institutions. If the different approaches produce identical rankings of the prisons examined, with the multilevel models providing the “gold standard,” then the correlation (Spearman’s rho) would be one. On the other hand, if there is absolutely no correspondence between the different approaches, then the correlation would be zero. While we examined one month only, any and all months could be examined with this same strategy.

In the multilevel models used here, the measure of performance is the Empirical Bayes (EB) residual, which is a statistically adjusted estimate of how much above (positive value) or below (negative value) the observed value falls from what was expected. For the measures of inmate misconduct, the EB residuals measured how much each prison either raised or lowered the probability that an inmate engaged in misconduct. Since misconduct was measured as a simple yes/no dichotomy, the multilevel models were modeled with an appropriate nonlinear specification (for details, see Camp et al., 2003).

As a first test of the consequences of ignoring individual-level information, complete multilevel models for June 2001 were compared to specifications that did not include individual-level variables. The new models included all of the same variables as the complete models, but the individual-level variables were aggregated to the organizational level.¹¹ The rankings of the prisons on the empirical Bayes residuals, the organizational performance measures, derived from the respective models were then compared for seven different types of misconduct: all forms of misconduct, violent forms of misconduct, drug-related misconduct, misconduct related to the security of the prison, property-types

¹¹There were slight deviations from models previously published and those discussed here (Camp et al., 1998). For example, including the sex of the inmates caused estimation problems for the models of some dependent variables and was dropped from those models. For the same reason, the squared terms for age and base score that appeared previously were dropped.

of misconduct, misconduct related to inmate accountability (being in the proper place, etc.), and other types of misconduct not otherwise classified.¹² Actually, all forms of misconduct is an overall category capturing all of the other forms of misconduct. The other six categories of misconduct parcel the all category into constituent parts. The results of comparing the rankings are presented in Table IV.1.

The performance measures derived from the complete models, with both individual- and organizational-level variables, and the reduced models, with only organizational-level variables, produced similar rankings of the 156 prisons examined in June of 2001. A perfect correspondence between the measures would have generated a correlation of 1, and as can be seen in Table IV.1, almost all of the correlations were at 0.9 or greater. The one exception was the correlation between the complete and reduced models for drug-related misconduct, but the correlation was greater than 0.89, which is virtually indistinguishable from 0.90. While the complete models would still be preferred on a statistical basis, it does not appear that large amounts of non-reliability are introduced into the rankings of the prisons if the reduced models with only the organizational-level variables provide the performance measures. Of course, even in the aggregate-only multilevel models, individual propensity to engage in misconduct is the outcome variable, not the count of instances of misconduct at each prison as in the regression models to which we now turn.

The second test of relying upon data only collected at the organizational level was to compare the multilevel results for complete data models with ordinary regression results conducted on organization-only data. For the regression models of the aggregate data, the outcome variable was the actual count of the number of instances of misconduct at each particular prison.¹³ Since the outcome variable was a count variable, the data were most appropriately modeled with a negative binomial distribution (Long, 1997: Chapter 8). The residual value analyzed in this case was based on the difference between the actual

¹²The BOP has a fairly detailed list of infraction codes that are used in the inmate disciplinary process. It is necessary to group the codes in some manner to produce sufficient monthly counts that are amenable to statistical modeling. A list of the BOP codes and the corresponding groupings is presented in Appendix 1.

¹³More detail about the specification of the negative binomial regression models is presented in the subsection titled "Aggregate-Only Models of Institutional Performance."

number of instances of misconduct at the given facility and the expected number of instances given the understanding built into the regression model. As such, the residuals for the multilevel and negative binomial regression approaches are in different units, but we are most interested in the respective rankings of prisons produced by the two approaches and not the actual values of the residuals. As much as possible, the variables analyzed in the aggregate-only models matched the individual-level characteristics found to be important in the multilevel analysis reported in *Justice Quarterly* (Camp et al., 2003) and are similar to those used in the aggregate-only multilevel models discussed in Table IV.1.

The results comparing the performance measures created from the aggregate-only models and the complete multilevel models are presented in Table IV.2. Correlations were calculated between the empirical Bayes residuals for multilevel models and two different residuals for the count models of the aggregate data. Both the Pearson and the standardized Pearson were examined for the aggregate count models. Since the variance of the residual is a function of the mean in a negative binomial model, a raw residual is not appropriate as institutions with large counts are more likely to have large residuals simply because the count was larger (Zelterman, 2002: 38-39). As such, either the Pearson residual or the standardized Pearson residual gives a more accurate depiction of the deviance from the expected value. As can be seen in Table IV.2, using one or the other of the Pearson and standardized Pearson does not have much practical impact upon the results.

Only three of the possible seven categorizations of misconduct were examined for these purposes. All misconduct was found in prior analyses to be the most reliable form for constructing performance measures and was examined for that reason. Violent misconduct and drug misconduct, although not as reliable in producing performance measures, were examined because of their substantive importance. These three outcome variables—all misconduct, drug misconduct, and violent misconduct—are also examined later to compare the results of longitudinal models of misconduct. However, the section on performance measures derived from aggregate-only data relies more upon an alternative categorization of misconduct that is of importance to the BOP. That categorization is based upon the severity level implied in the misconduct codes used by the BOP. In that regard, 100-level offenses are the most serious, followed by 200-, 300- and 400-level offenses. As such, all 100-level offenses are categorized together, all 200-

level offenses are grouped together, but 300- and 400-level offenses are considered together as most BOP staff do not differentiate between the seriousness of a 300- and a 400-level charge.

The rankings of the institutions as determined by the multilevel and aggregate count models do correlate at a fairly high level. For the outcome variables examined, the correlation between the two methods of constructing performance measures for all misconduct was 0.75, which is acceptable if not ideal. For violent misconduct and drug misconduct, the correlations were slightly more modest at about 0.64 and 0.67 respectively. This means that there is considerable overlap in whether the institutions ranked high or low based upon the respective methods. However, the rankings were not identical, as identical rankings would have produced correlations of 1.0, and this was clearly not the case. Another way of thinking about the commonality in rankings is by calculating how much of the variance is shared in the rankings whether based on the multilevel or aggregate-only approach. This can be calculated by squaring the correlation between the two sets of rankings. For all misconduct, this value (based on the Pearson residual) is 0.570, for drug misconduct 0.444, and for violent misconduct 0.413. In other words, about 57.0 percent of the variance in ranking the prisons based upon all misconduct is shared between the multilevel and aggregate approaches. The shared variance for drug misconduct is 44.4 percent, and 41.3 percent of the variance is shared between the two methods when the outcome variable is violent misconduct. While these are respectable amounts of shared variance by normal social science standards, there is still much variance that is not accounted for when the different methods are employed.

All other things being equal, the multilevel models of institutional performance are preferred over the aggregate-only models, especially for drug and violent misconduct. For all misconduct, the variance shared between the two approaches of measuring prison performance was almost 60 percent, but for drug and violent misconduct, the shared variance was below 50 percent. For the practical reasons enumerated above, we proceed with aggregate-only analyses of institutional performance. For the sake of completeness, we start by presenting comparisons of TCI and BOP prisons based upon complete multilevel models, at least for a portion of the time period examined.

Table IV.1
Correlation between Complete and Incomplete Multilevel Specifications of Inmate Misconduct*

Type of Misconduct	Spearman's Correlation	Probability of Correlation
All forms of misconduct	0.906	0.000
Violent misconduct	0.905	0.000
Drug-related misconduct	0.891	0.000
Property misconduct	0.919	0.000
Accountability misconduct	0.953	0.000
Security misconduct	0.969	0.000
Other misconduct	0.901	0.000

* The complete specification of the models of inmate misconduct included the following variables at the individual level: base custody score, age, count of prior instances of misconduct, amount of time at risk of misconduct, citizenship (Mexican, Cuban, and other with U.S. being the reference group), race (black and other with white being the reference group), and Hispanic ethnicity. Sex was not included as once the other covariates are included in the model, sex is not significant in any of the models examined and causes estimation problems for some dependent variables. The complete model had the following variables at the aggregate (organizational) level: average age, crowding, an integration index of the racial groups, and security level (minimum, low, medium, and administrative with high security the reference group). The reduced model (aggregate only) did not contain any individual-level variables and had the following aggregate measures: average custody score, average age, percentage of inmates who are black, percentage of inmates who are other race than black or white, percentage of inmates who are Hispanic, percentage of inmates who are Mexican, percentage of inmates who are Cuban, percentage of inmates who are of some other nationality than U.S., Mexican, or Cuban, crowding, and security level (minimum, low, medium, and administrative with high security the reference group). The integration index was dropped as it is comprised of the same variables that were used to capture the percentages of the different racial groups.

Table IV.2
Rank Correlation between Individual-Level (Multilevel) and Aggregate Count
Specifications of Inmate Misconduct*

Type of Misconduct	EB and Pearson Residual	EB and Standardized Pearson Residual
All forms of misconduct	0.755**	0.751**
Violent misconduct	0.643**	0.643**
Drug-related misconduct	0.667**	0.668**

* The multilevel, individual-level specification of inmate misconduct included the following variables at the individual level: base custody score, age, count of prior instances of misconduct, amount of time at risk of misconduct, citizenship (Mexican, Cuban, and other with U.S. being the reference group), race (black and other with white being the reference group), and Hispanic ethnicity. The multilevel model had the following variables at the aggregate (organizational) level: average age, crowding, percentage of inmates who were black, percentage of inmates who were of other race than black or white, percentage of inmates who were Hispanic, whether the prison was primarily a female prison, and security level (minimum, low, medium, and administrative with high security the reference group). The aggregate specification of the count of the instances of misconduct tried to mimic the multilevel specification where possible and included variables for average age of inmates, crowding, percentages of inmates who were black, percentage of inmates who were of other race than black or white, percentage of inmates who were Hispanic, whether the prison was primarily a female prison, and security level (minimum, low, medium, and administrative with high security the reference group).

** Correlation significant at $p \leq 0.001$.

Multilevel Models of Institutional Performance

The analysis presented in this section extends the previous work on inmate misconduct conducted by ORE researchers (Camp et al., 2003) in an important and interesting way. The earlier analysis examined all of the BOP prisons operating in June of 2001 to provide a cross-sectional snapshot of performance at that point in time. The current analysis extends that framework to develop cross-sectional views for a greater number of points in time. The analysis extends from January 1999 to December 2001. Even though TCI began receiving inmates at the end of 1997, the starting point of January 1999 was

selected to allow for initial operations to have settled into a routine at TCI before starting a formal comparison. The data end in 2001 because of the difficulty in putting together the data at the individual-level data.¹⁴

The comparisons of the different low-security prisons were made with graphs derived from the results of multilevel models estimated with the HLM 5.0 software (Raudenbush, Bryk, & Congdon, 2001). For each of the 36 months examined, models for the seven different categorizations of misconduct were specified in the HLM software. After estimating the equations, residual files were generated for each of the analyses that provided monthly performance measures for each type of misconduct. The results from the seven types of misconduct, each over 36 months, were joined together in seven computer files, and the graphs for each respective type of misconduct were produced using PROC GPLOT in SAS (SAS International, 2002). In addition to plotting the actual Empirical Bayes performance measures for each month, the SAS software was used to estimate and plot a regression line to represent the trend in misconduct for the different prisons over the entire time span. The regression model chosen to represent the trends in misconduct included a square and cube term that allowed the lines to bend and change direction.¹⁵

Three different views of the data are presented for each type of misconduct. While somewhat redundant, the views do provide different types of information about the trends in misconduct. In the first representation, only the regression lines are presented for each of the low-security prisons in the BOP and TCI. The HLM models of performance were actually based upon all BOP facilities (all security levels), but a graph of all prisons is too busy to allow for differentiation of individual prisons. The first view of the data allows for examination of how TCI and the three BOP comparison prisons performed with respect to other BOP low-security prisons. Primarily, these graphs are presented to allay

¹⁴Due to changes in personnel and misplaced computer programs in the Office of Research and Evaluation, it was not possible to simply extend the data collection from the end of 2001 until the end of 2002 or into 2003. Instead, the programs, which are complicated, were recreated from scratch, and their completion did not occur in time to allow for the significant data testing necessary to insure the quality of the data.

¹⁵The equation is $Y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + e_i$ where i indexes each institution, x represents month, and e is the normal error term.

any concern that the comparisons between TCI and the three BOP comparison prisons are stacked in any manner against TCI. The second view of the data presents the regression lines for only TCI and the three BOP comparison prisons. These graphs also include the actual monthly performance values for each of the prisons. Since these graphs are less cluttered than the first representation, it is easier to determine the relative performance of TCI and the three BOP comparison prisons. Finally, the last representation of the data again presents monthly performance values and the regression line, but only for TCI. In addition, a 95 percent confidence interval around the estimated regression is presented. This allows for an inspection of whether the different parts of the estimated trend include a value of zero. A value of zero, which is represented by a horizontal reference line in all of the different graphs of misconduct, indicates the expected value for each institution. Values above the reference line indicate that the institution contributed to inmate misconduct whereas values below the reference line mean that the institution detracted from the probability that an inmate would be involved in misconduct. In other words, if the confidence interval around the regression includes zero, it is not possible to conclude that TCI is performing differently than would be expected given the characteristics of the inmates there and the institution itself, as captured in the HLM models.

The first graphs generated from the multilevel models are for all types of misconduct considered together. As noted from previous analysis (Camp et al., 2003), all misconduct is the most reliable form of misconduct for generating performance measures to compare prisons. The first view of all misconduct which includes all low-security prisons is presented in Figure IV.1. As can be seen there, TCI contributed to a higher probability that inmates would be involved in overall misconduct for much of the time period than any of the BOP comparison prisons. Taft started off with very favorable negative values for the first six months of the observation period, but after that initial period all of the observations are above the expected level although there is a trend downward toward the end of 2001. Prior to the downward trend, though, TCI was one of the worst performing low-security prisons.

{Figures IV.1, IV.2, and IV.3 about here.}

The three BOP comparison prisons showed different performance than TCI. The probability of misconduct at FCI Yazoo City was pushed to some of the lowest levels for any of the BOP low-security prisons. For FCI Elkton and FCI Forrest City, the

performance was more modest as they performed in the middle of the pack. Nonetheless, as the values at these institutions hovered around zero, the institutions contributed to inmate misconduct in a manner not very different than would have been expected otherwise. It is interesting to note, though, that TCI had a more favorable effect upon the probability of inmate misconduct for the first couple of months of the observation period.

The trends noted from Figure IV.1 are reinforced with the information contained in Figures IV.2 and IV.3. Whereas the performance measures for TCI were initially below the zero reference line, they quickly moved above the line and stayed there for the rest of the observation period. The values for FCI Yazoo City, on the other hand, were consistently below the zero reference line with few exceptions. The confidence interval presented for the regression line for TCI in Figure IV.3 reinforces the idea that performance at TCI was less than desirable for most of the observation period. Performance was good early on as noted previously, but the regression line moved above the zero reference line early on and remained there. The confidence interval shows that until the very end of 2001, the performance was significantly worse than expected for most of the observed time period. Clearly, though, performance was moving in a favorable direction (toward 9) by the end of 2001.

{Figures IV.4, IV.5, and IV.6 about here.}

Information on violent forms of misconduct are presented in Figures IV.4, IV.5, and IV.6. The information provided in these graphs suggests that whatever the problems at TCI that contributed to negative performance ratings for all types of misconduct considered together, the problems are not related to violent forms of misconduct. As can be observed in Figure IV.4, being an inmate at TCI reduced the probability of violent misconduct during the observation period. In fact, TCI was one of the better performing prisons of all low-security prisons during this period. The performance at FCI Forrest City and FCI Yazoo City was also favorable for most of the observation period, as inmates at these institutions had lower probabilities of violent misconduct, although FCI Forrest City started the observation period with higher contributions to violent misconduct. Performance at FCI Elkton, though, was in marked contrast. For almost the entire observation period, inmates at FCI Elkton were more likely to be involved in violent misconduct than expected, although this negative performance trailed off at the end of 2001.

The ability to compare TCI to the BOP prisons is enhanced in Figure IV.5. Whereas performance at TCI, FCI Forrest City and FCI Yazoo City did not differ in a major fashion, performance at FCI Elkton was substantively different than the performance at these other prisons. One thing to note from both Figures IV.5. and IV.6 is the greater dispersion of performance measures from one month to the next. For example, even though the performance of FCI Elkton was generally negative in that inmates there were more likely to be involved in misconduct, not all of the monthly observations for FCI Elkton fell above the zero reference line. Another reflection of the greater dispersion of values from month to the month is the much wider confidence interval noted for TCI in Figure IV.6. Where the information presented in Figures IV.4 and IV.5 suggested that inmates at TCI were less likely to be involved in misconduct, the information in Figure IV.6 shows that statistically the performance did not differ from the expected value as zero falls within the upper confidence interval for almost every month. In fact, although the needed graph is not presented here, the performance at TCI and FCI Elkton only differed in a statistically significant manner for several of the months during the observation period when the confidence intervals around the estimated regression lines are considered for both institutions. This emphasizes both the need to have reliable measures and to consider reliability when making comparisons. Otherwise, erroneous conclusions can be reached.

{Figures IV.7, IV.8, and IV.9 about here.}

Information about drug misconduct at the respective prisons is presented in Figures IV.7, IV.8, and IV.9. Part of the problem experienced at TCI with all forms of misconduct considered together was related to the impact of TCI upon inmate drug misconduct. As can be seen in Figure IV.7, TCI had the highest contribution to the probability of drug misconduct of all low-security prisons examined for most of the time period. Almost as low in terms of performance was the contribution of FCI Forrest City to drug misconduct, although the poor performance of both FCI Forrest City and TCI did tail off dramatically at the end of 2001.

As is most clear in Figure IV.8, performance at FCI Elkton and FCI Yazoo City started at very similar levels as that at TCI and FCI Forrest City. However, performance improved to where the lines fell below the zero reference line for part of the time period observed. Nonetheless, performance became worse at the end of the observation period, and FCI

Elkton actually performed worse than TCI and FCI Forrest City for a few months at the end of the observation period.

The information in Figure IV.9 shows that the confidence interval around the projected regression line was again larger than observed for all forms of misconduct considered together. Again, this is a function of the lower reliability of basing performance upon only drug misconduct. Nonetheless, performance at TCI pushed the probability of inmate drug misconduct so far above zero that for most of the time period observed that the performance at TCI was worse than expected.

The graphs for the other forms of misconduct considered with multilevel models—accountability, security, property, and other—are presented in Appendix 2. These forms of misconduct are no less important than those considered so far to prison administrators, but they are not as often the focus of scrutiny by those outside of corrections. Generally speaking, the results presented in these graphs demonstrate that TCI did worse than the comparison prisons for these types of misconduct, with the exception of security misconduct where the performance of TCI was very good. For accountability misconduct, property misconduct, and other types of misconduct, the performance of TCI was worse than expected for most of the observation period.

In sum, the performance at TCI as measured by the multilevel models of individual-level data was generally less favorable than the performance of the BOP comparison prisons. As a general rule, none of the four prisons were stellar performers when compared to other low-security prisons in the BOP. Taft did perform well on two categories of misconduct, violent misconduct and security misconduct, but this favorable performance was not sufficient to offset the poor performance in other areas of misconduct, as captured by the overall category of misconduct.

Aggregate-Only Models of Institutional Performance

The models of organizational performance, as derived from data at the institutional level only (or aggregate-only data), demonstrated very similar results to those presented above. The graphs of organizational performance upon all types of misconduct are presented in Figures IV.10, IV.11, and IV.12. The models included the aggregate variables used to compare the multilevel models with the aggregate-only models as discussed above. In

addition, several other variables were included in the models to improve model fit. The models of misconduct included independent variables for average age of the inmates, percentage of inmates who were black, percentage of inmates who were of some other race than white or black, percentage of inmates who were Mexican citizens, percentage of inmates who were Cuban citizens, percentage of inmates who were of some other citizenship than U.S., Mexico, or Cuba, whether the prison housed males or females, percentage of inmates who were Hispanic, crowding level at the prison, number of inmates at the prison, percentage of inmates who had instant offenses that were of either high or greatest severity (as determined by BOP inmate classification system), percentage of inmates with highest levels of violence (as determined by BOP inmate classification system), percentage of inmates convicted of various types of offenses (drug/liquor, firearms/explosives, violent, property, white collar, immigration, extortion/fraud/bribery, sex, and robbery constituted the separate categories), median length of sentence, percentage of inmates convicted of offenses in Washington, D.C., percentage of inmates in gangs classified as strategic threat groups by the BOP, percentage of inmates eligible for furlough, and security level of the prison.

{Figures IV.10, IV.11, and IV.12 about here.}

The results presented in the graphs for aggregate-only models incorporate a longer time span, from January 1998 to June 2003. As can be seen in Figure IV.10, when looking at all misconduct, TCI started out as one of the better performing prisons, but by the end of the time period the counts of all types of misconduct were higher than expected. TCI had one of the worst records in this regard from about early 2000 until the end of the time period.

The uneven performance of TCI is better illustrated in Figure IV.11. In that graph, it is obvious that the observed counts of misconduct were lower than expected (the negative residuals), but that pattern ended in early 1999 where the counts of all types of misconduct became greater than expected (positive residuals). It should also be noted that two additional reference lines show up for the first time in Figure IV.11. The observations between the two perpendicular reference lines cover the same time period used for the multilevel models, e.g., January 1999 to December 2001. These lines are added to the graphs to assist in the visual comparison of the graphs derived from multilevel models and aggregate-only models. Comparing the respective portion of

Figure IV.11 back with Figure IV.2 (which was for all misconduct as derived from multilevel models), it is easy to see that the same pattern of performance is told by performance measures derived from both methods. There are some slight differences, but TCI is depicted by both methods as being one of the poorest performing low-security prisons.

The information in Figure IV.12 confirms that TCI differed from the expected value for BOP prisons. The confidence interval for the regression line drawn through the monthly measures does not include zero (expected level of performance) except for a very few months. These exceptions occurred as performance at TCI led to higher counts of all types of misconduct than expected.

For comparative purposes with the multilevel results presented earlier, performance measures for the break-outs of violent misconduct and drug misconduct are presented here. However, the major emphasis of this section is on partitioning the information provided by all types of misconduct into categorizations of misconduct based upon the BOP charging system. Charges in the BOP range from 100-level offenses, the most serious, to 200-level charges, and 300 and 400 level charges. The real demarcation is between 100-level offenses, 200-level offenses, and 300-level and 400-level offenses combined, and the outcomes were categorized that way for the analyses presented below.

{Figures IV.13, IV.14, and IV.15 about here.}

Figure IV.13 shows the plot of the trends in performance measures derived from violent forms of misconduct in the BOP using aggregate data. As is suggested there, the counts of this type of misconduct are at about the expected level for TCI as most of the monthly measures fall around zero. It does look as though performance at FCI Elkton may be problematic for violent misconduct as the counts are higher than expected, although counts of violent misconduct fall to lower than expected values at the end of the time period.

A more detailed comparison of TCI and the BOP comparison prisons is presented in Figure IV.14. As can be seen there and in Figure IV.13, most of the lines representing trends in violent misconduct actually fall close to one another. The same type of

performance for Taft and the three comparison prisons was noted when the performance measures were based upon multilevel models as well (see Figure IV.5).

Figure IV.15, which places a 95 percent confidence interval around the regression line through the monthly measures for TCI, demonstrates that performance at TCI was close to expected levels, at least for the trend. The confidence interval includes the value of zero for all time points. However, there were some months when performance at TCI deviated from expected levels in a substantial manner. The monthly points in the graphs of models based upon aggregate-only data are standardized. Generally speaking, any standardized residual larger than 2 or smaller than -2 are considered to be outliers. As can be seen in the graph, TCI had such values that indicated significantly higher counts of violent misconduct than expected (monthly values ≥ 2) as well as values significantly lower than expected counts (monthly values ≤ -2). Further attention is given to interpreting monthly performance later in this portion of the report.

{Figure IV.16, IV.17, and IV.18 about here.}

Figure IV.16 shows that drug misconduct was a problem at TCI as the counts are generally higher than expected (regression line above 0) after an initial period where the counts were less than expected (regression line below 0). By the end of the time period, TCI was performing at a lower level on this measure than any BOP low-security prison. FCI Forrest City also had a problem with higher counts of drug misconduct than expected, but performance on this measure turned around at FCI Forrest City toward the end of 2002 and dropped to acceptable levels (around the expected deviation value of 0). The trend at TCI though, was for drug misconduct to increase through the end of 2002 and into 2003.

The trends for TCI and the comparison prisons are more easily seen in Figure IV.17. The results of the aggregate-only model of drug misconduct, as summarized in Figure IV.17, provide slightly different information than was given by the multilevel models (see Figure IV.8). Where the multilevel models suggested that performance at TCI was improving at the end of 2001 and counts were moving toward the expected levels, Figure IV.16 demonstrates that performance was at a constant for most of 2001 at a level higher than expected. In 2002 and through June 2003, the counts of drug misconduct at TCI continued to climb to new levels above expected counts. Also, Figure IV.17 shows that

where performance at FCI Yazoo City was acceptable for most of the time period, performance was problematic at the end of the time period as counts are at higher levels than expected for much of 2002 and the first six months of 2003.

Figure IV.18, which places a confidence interval around the trend line, confirms that performance at TCI was problematic. The trend line deviates from 0 for most months, and is well above 0 for almost all of the latter months of the time period depicted.

As noted previously, performance at TCI was problematic when all types of misconduct were considered together. In other words, counts of all misconduct were higher than expected. The rest of this portion of the report tries to determine whether these higher than expected counts were consistent across the board for all levels of misconduct severity, or whether the counts existed for only certain types of misconduct, again, as classified by severity. Three categories of misconduct were examined as noted previously. The offenses go from the most serious (100-level offenses), to the serious (200-level offenses), to the least serious (300 and 400-level offenses grouped together).

{Figures IV.19, IV.20, and IV.21 about here.}

The results in Figure IV.19 demonstrate that TCI had a problem with higher counts than expected for 100-level offenses during most of the time period between 1998 and the first half of 2003. The trend reflected in the regression line suggests that counts were initially lower than expected at TCI, but the counts quickly moved to levels that were indicative of problems. In fact, from around June 2001, the trend line for TCI is much higher than that for any other BOP low-security prison.

Results summarized in Figure IV.20 demonstrate that trends for the BOP comparison institutions are less problematic in that they tend to hover close to 0 or even below 0 (especially FCI Yazoo City and FCI Elkton). The confidence interval for the trend line through the TCI monthly measures, presented in Figure IV.21, confirms that TCI was performing worse than expected (had higher counts of 100-level offenses than expected).

{Figures IV.22, IV.23, and IV.24 about here.}

The results incorporated into Figure IV.22 for 200-level misconduct for TCI and the comparison prisons are difficult to distinguish, with the exception of the performance of FCI Elkton. FCI Elkton started the time period with a trend indicating higher than expected counts for 200-level offenses, but the trend steadily declined over the period to where counts of 200-level offenses at FCI Elkton were lower than expected.

A better representation of the performance of TCI and the comparison prison is given in Figure IV.23. As can be seen in these results, performance at TCI for 200-level offenses was about that expected, at least with respect to the trend representation. The trend hovered around the expected value of 0, although there was a period during 1999 when counts were higher than expected. Figure IV.24 demonstrates that the confidence interval incorporated 0 for most of the time period, including most of the period during 1999.

{Figures IV.25, IV.26, and IV.27 about here.}

The results presented in Figure IV.25 demonstrate that TCI experienced problems with 300- and 400-level misconduct, at least for most of the time period. As was also noted for all types of misconduct and 100-level misconduct, performance at TCI was acceptable in the earliest months of the time period as the trend line falls below 0. However, from about April 1999 until the end of the time period, the trend line falls above 0. By the end of the time period, from about October 2000 onward, the difference from expected counts of misconduct is the highest of any of the low-security prisons.

The performance of the BOP comparison prisons is best seen in Figure IV.26. As can be interpreted from these results, the trends start at higher levels for the comparison prisons than for TCI. In fact, the comparison prisons all start with higher levels of this form of misconduct than expected. Quickly, though, expected values fell below 0 for all three comparison prisons and remained there throughout almost all of the observation period. With the exception of FCI Yazoo City, which does truly appear to have lower observed counts of 300- and 400-level misconduct, the expected values for the comparison prisons appear to be at about the expected level for the latter parts of the time period.

The results in Figure IV.27 confirm that the trend line for TCI deviates from the expected value of 0 for the latter portion of the time period. Also, it is easy to see in Figure IV.27

that many of the monthly measurements of performance fall at or above the 2 standard deviation threshold and that there is a general clustering of these high values for TCI.

There are two general statements to make about the aggregate-only model results. First, the models of performance based upon aggregate-only data appear to provide information about TCI and the comparison prisons that is similar to the results from the multilevel models. It does not appear that the aggregate-only models provide misleading information about comparative performance of TCI and the BOP comparison prisons. Second, the performance of TCI as reflected in the aggregate-only results was problematic. While these results only addressed issues related to comparable performance and not whether a baseline for acceptable performance exists, the results demonstrate that TCI compares unfavorably to both the BOP comparison prisons and most other BOP low-security prisons. The counts of all forms of misconduct considered together are higher than expected, and when broken up into constituent parts, it appears that the overall trend is influenced by problems with 100-level and 300- and 400-level misconduct. While not all forms of the substantive breakout of misconduct were examined, it appeared that part of the performance at Taft was related to drug misconduct.

Note on Interpretation of Residuals

The substantive interpretation of the residuals has not been emphasized in the discussion other than to note how the residuals from both multilevel models and negative binomial models (on aggregate-only data) can be used to discuss relative performance. The concern with relative performance springs from the purpose of the report, which is to assess the relative performance of TCI. Nonetheless, the monthly performance measures, the residuals, do have a substantive meaning as well.

Because we were comparing one prison to another in the section on aggregate-only models, for example, standardized residuals were used. The reason for this choice was the need to control for an undesirable feature of negative binomial models of counts. Namely, with count data, the absolute value of the residuals increase with larger counts independent of performance issues. For raw residuals, it would not be clear how much of the difference in values for two institutions was due to differences in performance and

how much was due to differences in the magnitude of the counts. Hence, there is a need for the standardized residuals.

When looking at just one institution, though, it is perfectly legitimate to use raw residuals, especially if the inspection includes one month at a time. A raw residual is simply the difference between the expected count for the institution and the observed count. The standardized residual for TCI in August of 2000, for example, was 1.81. A statistician knows that 33 percent of observations fall between the expected value of 0 and 1 standard deviation from 0. Another 14.5 percent of observations fall between 1 standard deviation and 1.96 standard deviations from 0. Since 50 percent of the observation fell below 0, the value of 1.81 meant that between 83 and 97.5 percent of the prisons had smaller deviations from the expected count of misconduct than was true for TCI. Admittedly, we could use a table to look up the exact percentage associated with a standard deviation of 1.81, but the interpretation still would not be that intuitive for managers.

While a standardized residual does not have intuitive meaning, and managers would have to be trained in their use, the value for the raw residual for Taft in August of 2000, 49.5, does have a more understood meaning. It means that TCI had 49.5 (the model produced expected values that are not integers) more instances of any form of misconduct than would have been expected in August of 2000.

Random Drug Test Results

The BOP has a program for administering random drug tests to inmates to discover the extent of drug use in BOP prisons. This is one of the few measures used by the BOP that is based upon a scientific methodology to uncover the underlying incidence of behavior. Models of the number of random positive test results were conducted using the same negative binomial method used for misconduct, although a number of the independent variables used for misconduct were dropped to insure model estimation.¹⁶

¹⁶The variables dropped included those measuring the percentages of inmates in the different categories of offense type, in security threat group gangs, in eligibility status for furloughs, and with Washington D.C. offenses.

As with any procedure, there can be deviations from the expected implementation. At TCI and several BOP prisons, the random drug tests were performed, and the results of the tests were obtained from a contract laboratory. However, instead of sending the drug samples under one account number for the low-security portion of the institution and a different account number for the minimum-security camp, all drug samples from TCI were sent using the same account number. This means that there was no means to separate drug test results from the low-security facility and the camp. Instead, the drug test results for TCI mixed the two facilities together, although the results were treated as though they were from the low-security prison only. If mixing the results from the two facilities created a bias, it would have favored the interests of TCI as minimum-security prisoners are lower security risks than low-security prisoners. Part of this risk includes severity of offense, including drug offenses. The same problem of using one account instead of two was noted for several BOP facilities.

Figure IV.28 presents the results of the model of positive results for random drug tests. As can be seen in the results, TCI experienced a higher number of positive results than would have been expected. In fact, for almost the entire time period examined, from late 1998 forward, TCI had a trend line indicating the highest number of observed values to expected values. All of the other low-security prisons were tightly grouped around the expected value, 0.

Figure IV. 29 shows that the three comparison prisons were very tightly grouped around the expected value of 0. Figure IV.30 shows that the trend line for positive test results at TCI did differ from 0, although the plot of the monthly measures shows that there were several months where the measures fell at or below the expected value of 0. Nonetheless, the regression line noting the trend was more heavily influenced by the greater number of monthly observations that were well above the expected deviation value of 0.

Misconduct at the TCI Minimum-Security Camp

All of the previous analyses, and most of the emphasis of this report, have focused upon the low-security (main) prison at Taft. There are several reasons for this, including the much larger size of the low-security facility (about three times as many inmates) and the availability of the low-security comparison prisons. Another reason for the focus upon the low-security portion is that the BOP has and continues to use private-sector

contractors to hold inmates in non-secure beds, in halfway houses that transition the inmate from prison to the free community. Although halfway houses are quite distinct in mission from minimum-security prisons, there is the clear similarity that both lack secure perimeters. Using private-sector contractors to hold inmates in secure beds was more of a conceptual leap and less common when the Taft contract was awarded.

Graphs representing the performance of the minimum-security camp at TCI are presented in Appendix 3 for the interested reader. Overall misconduct is presented as well as the breakout of all misconduct into 100-level offenses, 200-level offenses, and the combined 300- and 400-level offenses. There are no graphs comparing TCI to the three BOP comparison prisons because none of them were equipped with camp facilities that were comparable to that found at TCI. All of the comparison prisons eventually acquired camp facilities, but they were smaller than the one at TCI and did not operate over the same time span.

The graphs in Appendix 3 suggest that misconduct may have also been a problem at the Taft camp. The level of overall misconduct was higher than expected for most of the time period, although several BOP minimum-security prisons had higher deviations from expected counts. There was a troubling movement upward in the trend of all misconduct for TCI at the end of the time period (see Figures 3.1 and 3.2). An examination of the remaining figures in Appendix 3 show that the problem was not with 100-level offenses. The counts for the most serious of BOP offenses was lower at the Taft camp than expected. Instead, the higher than expected counts were noted for the 200-level offenses and the combined 300- and 400-level offenses. While the Taft camp was not in the most extreme portion of the graphs indicating poor performance for most of the performance measures, as was often the case for the low-security portion of the prison, the trends were not favorable for the most part.

Summary

The evidence presented in this section suggests that there were performance issues at TCI. TCI had higher counts than expected for most forms of misconduct, including all types of misconduct considered together. The results were consistent across the two methodologies examined here to create performance measures, the multilevel specification and the negative binomial models that used aggregate data only. The

aggregate-only models suggested that the problems in misconduct were centered around the most serious of BOP misconduct (100-level offenses) and the least serious (300- and 400-level offenses). TCI performed more along the lines expected for 200-level offenses.

Consistent with the problematic trends for misconduct at TCI, the results of the analysis of positive results for random drug tests also pointed to TCI as a potential source of trouble. TCI had the largest deviation of observed from expected values for most of the time period examined. The deviation was positive, meaning that TCI had more instances of positive findings for random drug tests than would have been expected.

Figure IV.1

All Misconduct at BOP Low-Security Prisons

Multilevel Models, January 1999 to December 2001

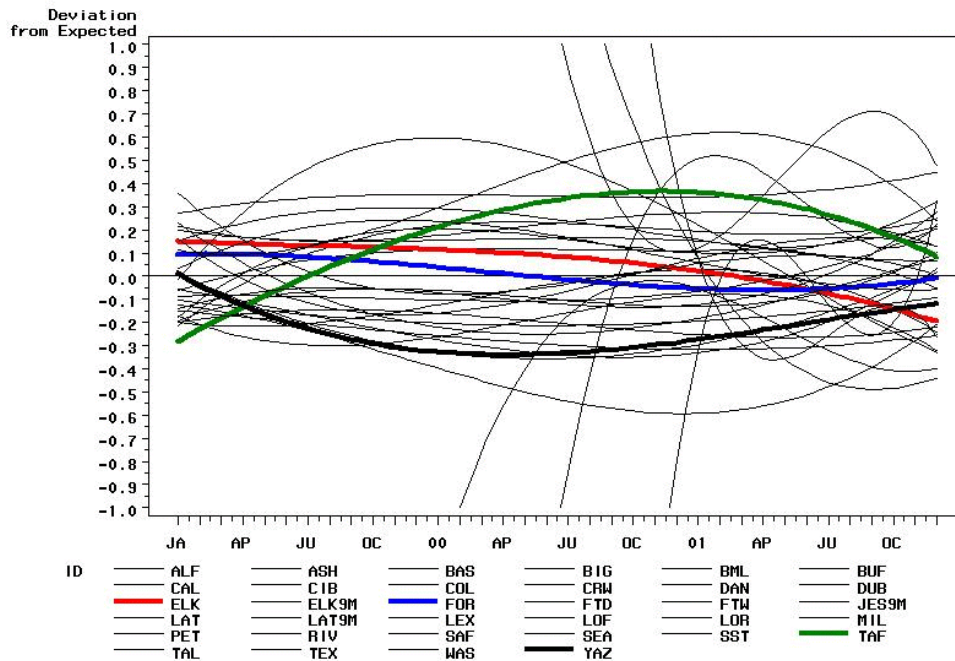


Figure IV.2

All Misconduct at TCI and Comparison Prisons

Multilevel Models, January 1999 to December 2001

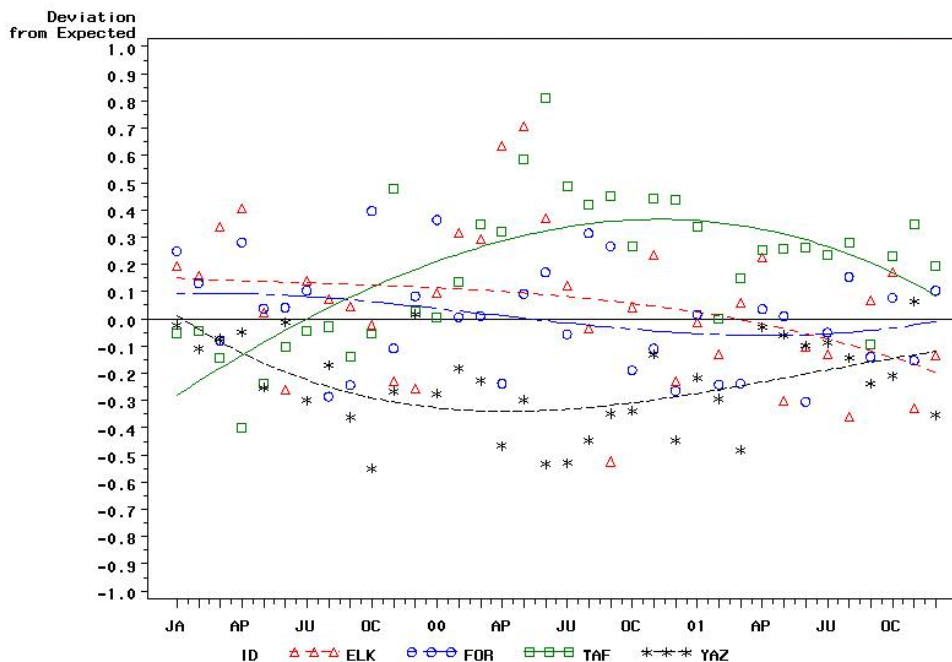


Figure IV.3

All Misconduct at TCI, with 95% Confidence Interval
Multilevel Models, January 1999 to December 2001

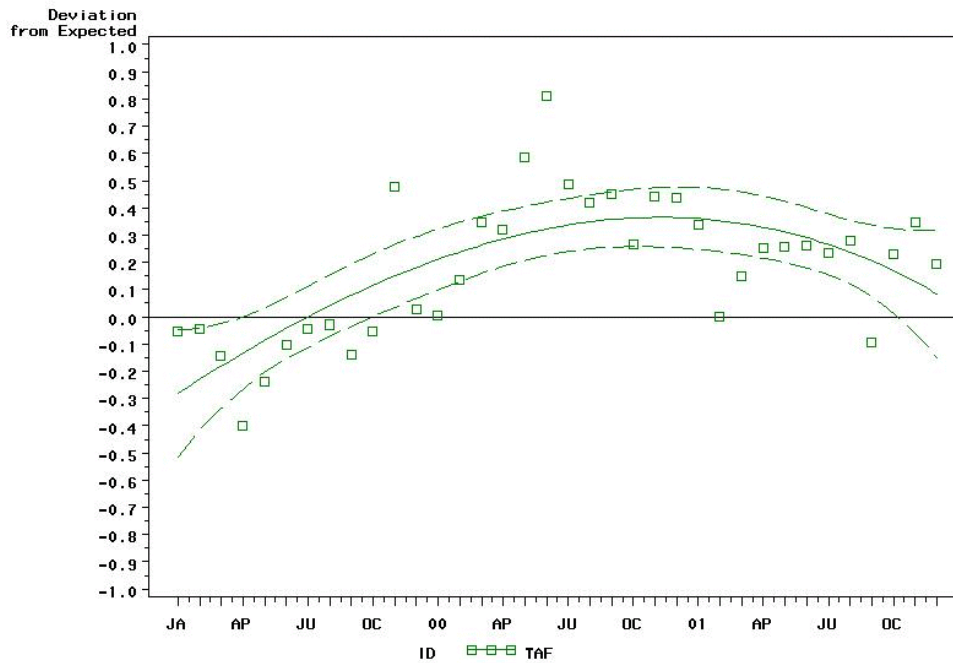


Figure IV.4

Violent Misconduct at BOP Low—Security Prisons
Multilevel Models, January 1999 to December 2001

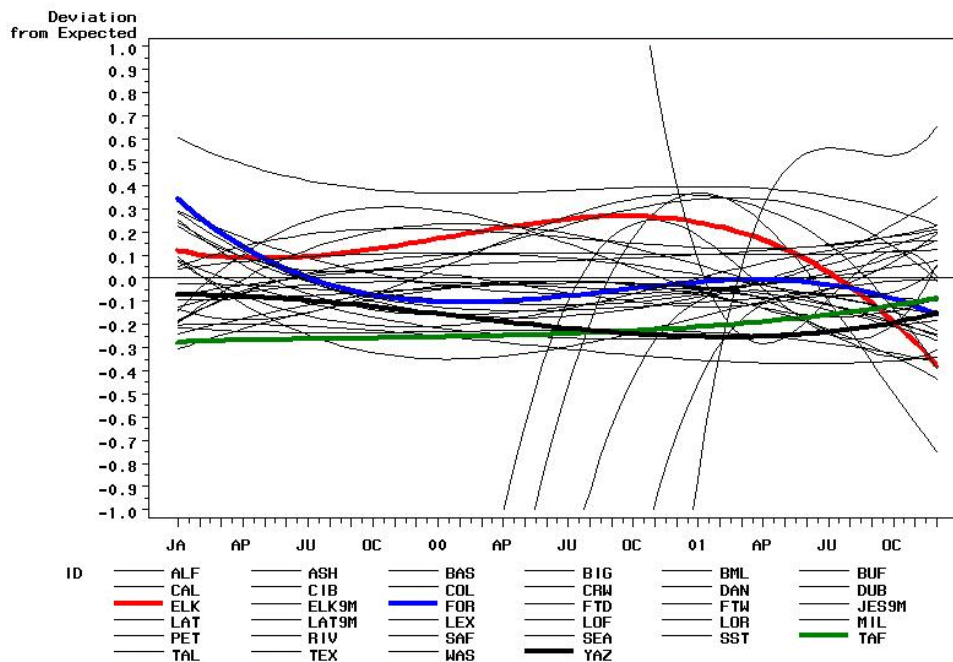


Figure IV.5

Violent Misconduct at TCI and Comparison Prisons

Multilevel Models, January 1999 to December 2001

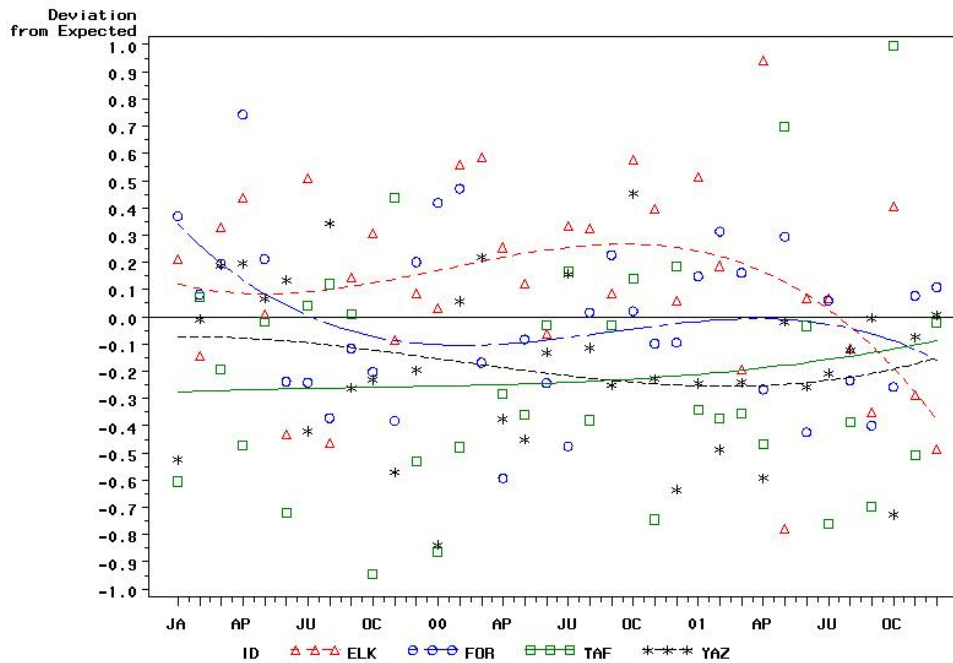


Figure IV.6

Violent Misconduct at TCI, with 95% Confidence Interval

Multilevel Models, January 1999 to December 2001

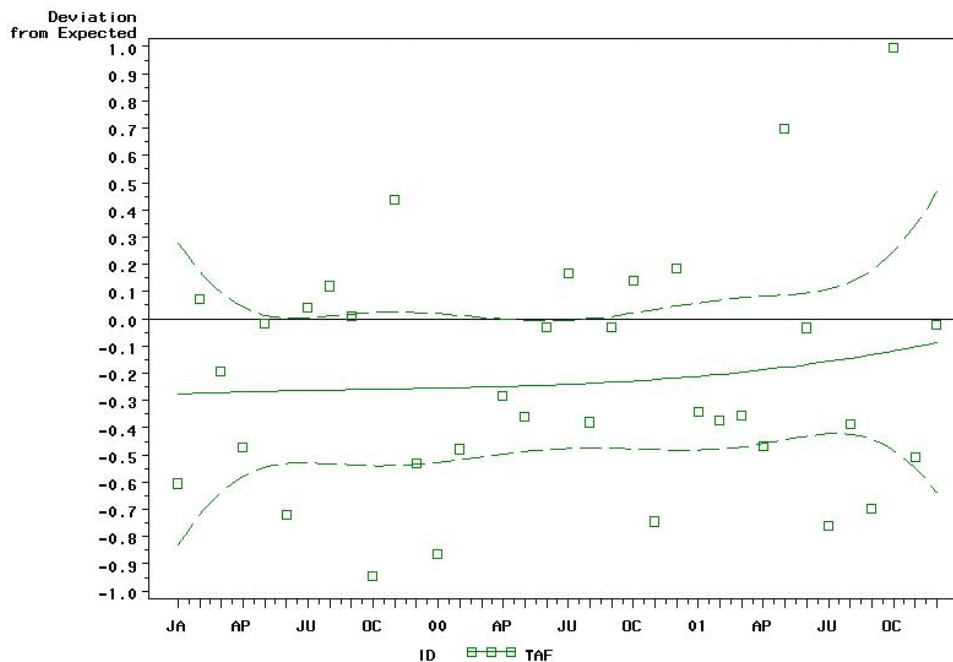


Figure IV.7

Drug Misconduct at BOP Low-Security Prisons

Multilevel Models, January 1999 to December 2001

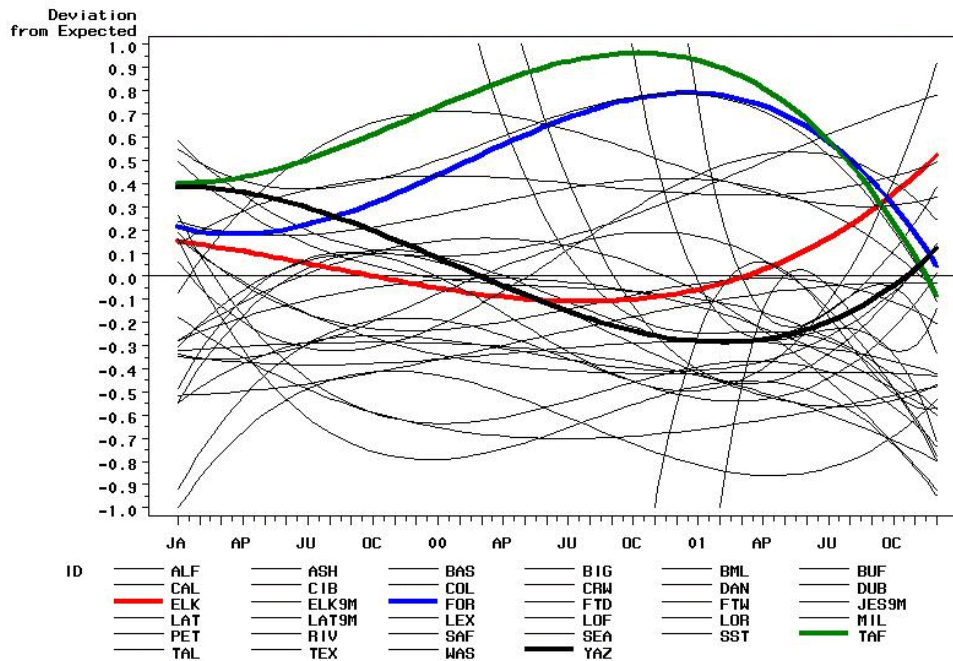


Figure IV.8

Drug Misconduct at TCI and Comparison Prisons

Multilevel Models, January 1999 to December 2001

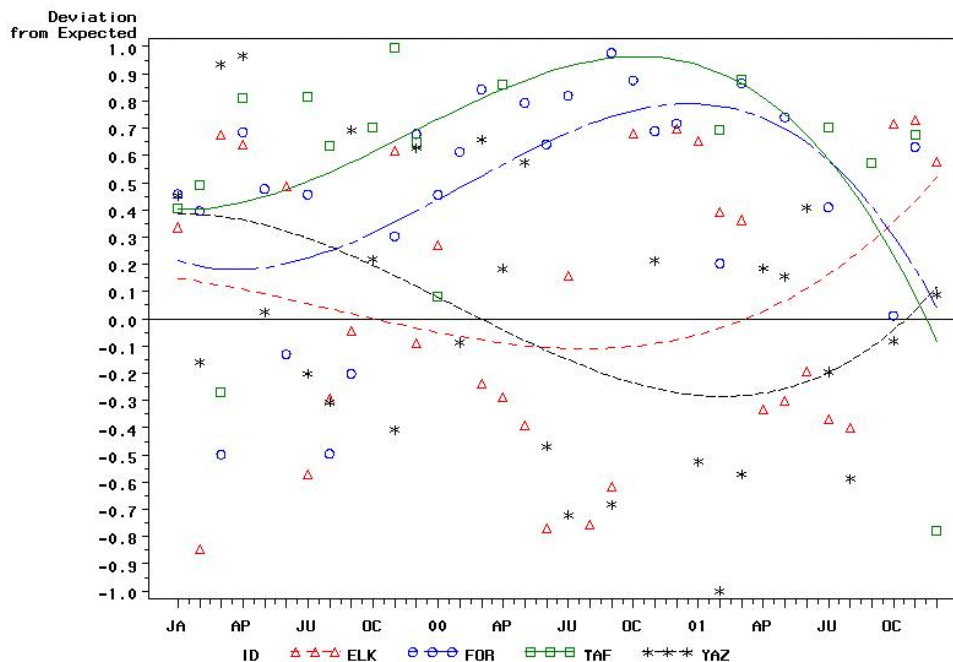


Figure IV.9

Drug Misconduct at TCI, with 95% Confidence Interval
Multilevel Models, January 1999 to December 2001

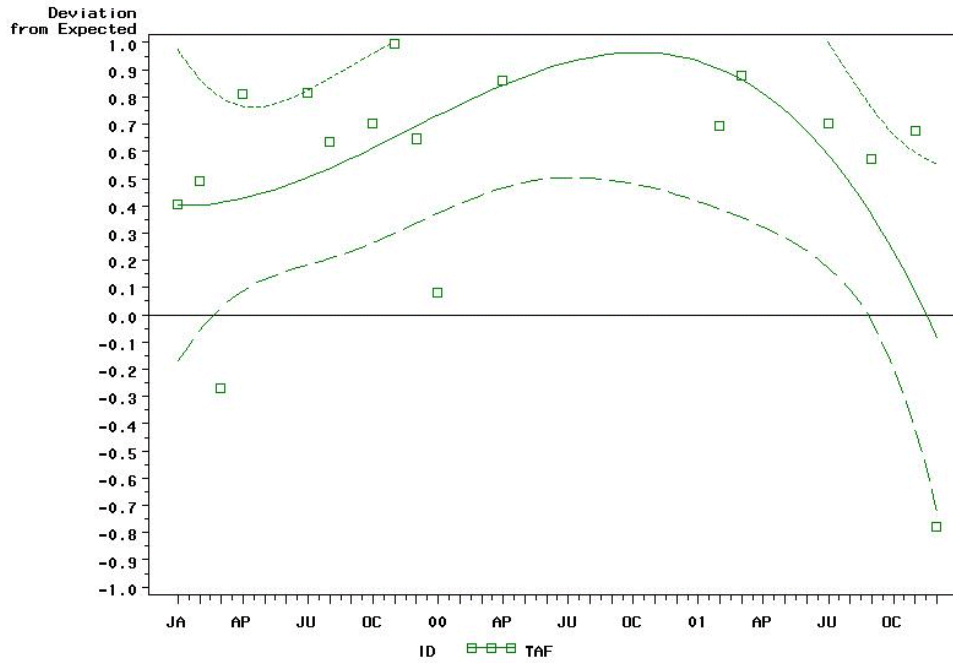


Figure IV.10

All Misconduct at BOP Low-Security Prisons

Aggregate Only Models, January 1998 to June 2003

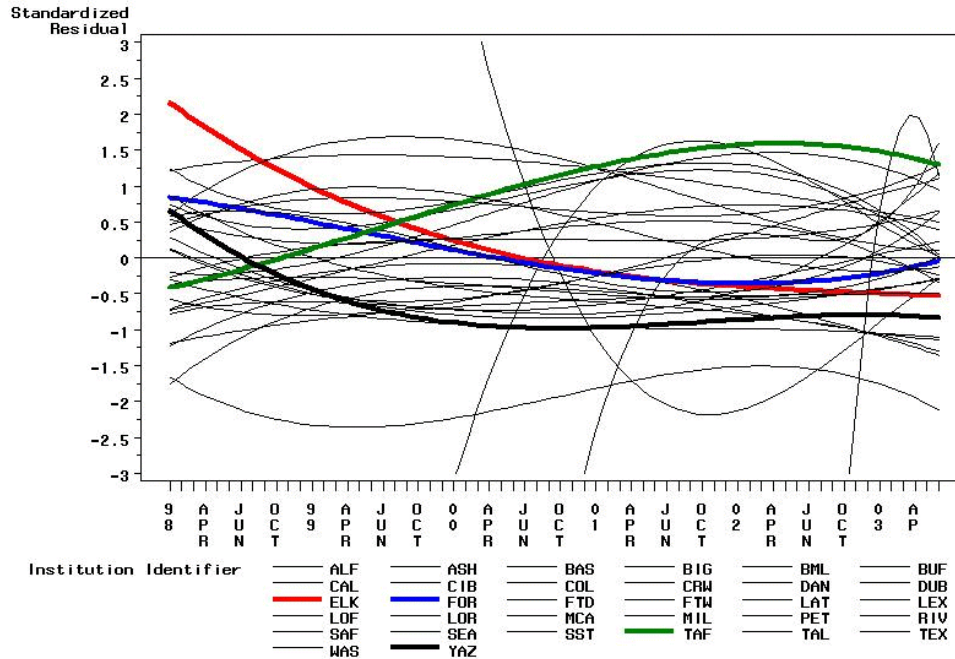
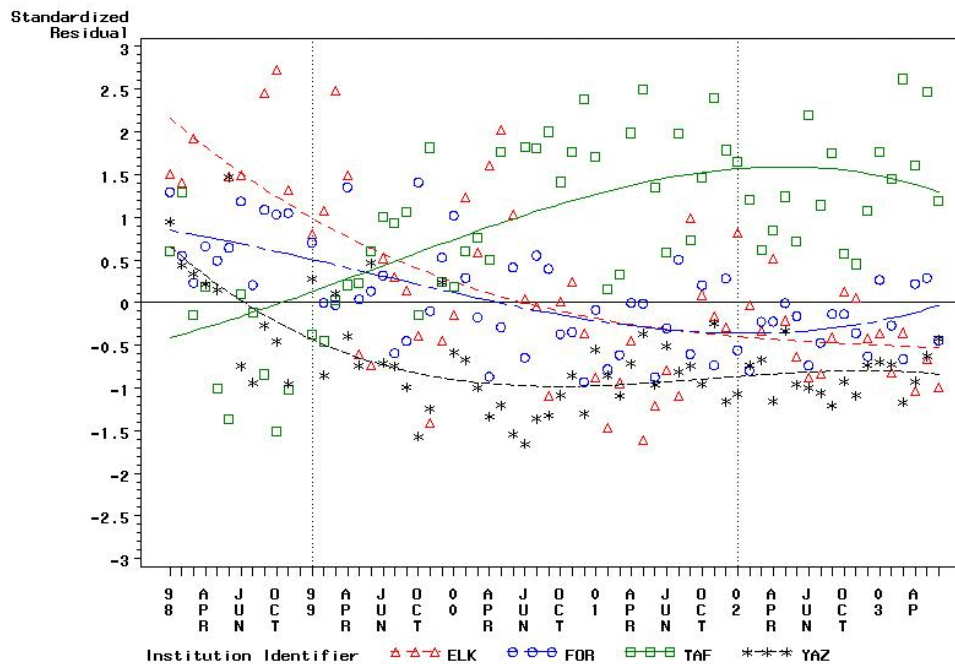


Figure IV.11

All Misconduct at TCI and Comparison Prisons

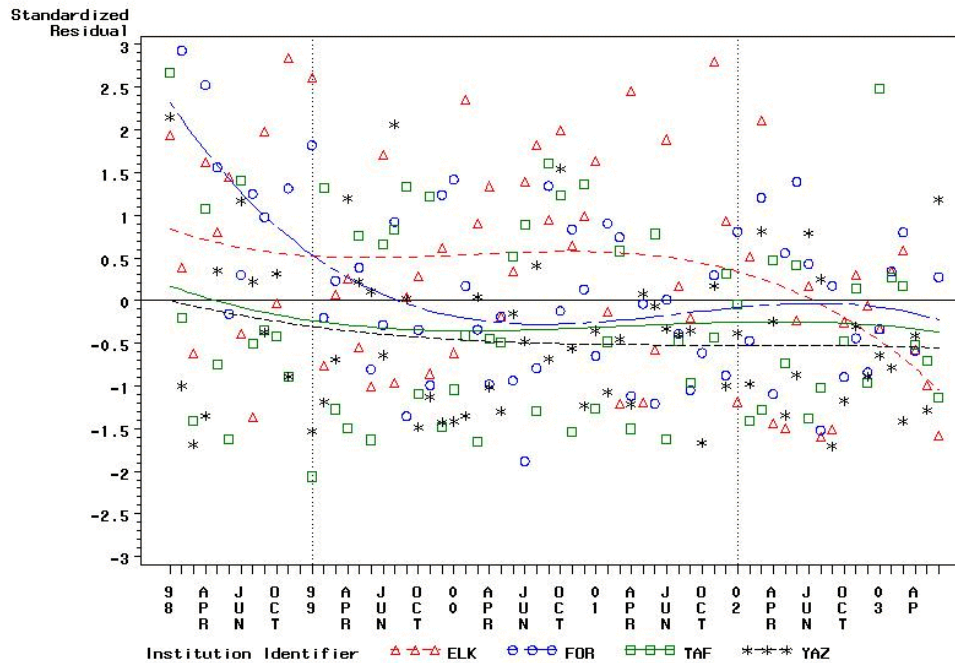
Aggregate Only Models, January 1998 to June 2003



All Misconduct at TCI, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003



Aggregate Only Models, January 1998 to June 2003



Aggregate Only Models, January 1998 to June 2003

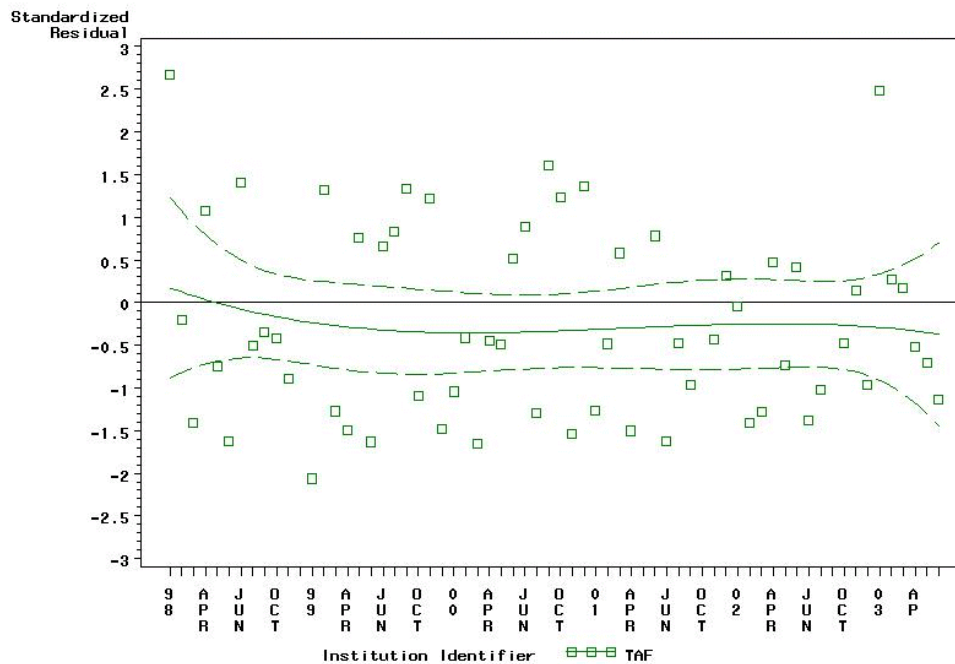


Figure IV.16

Drug Misconduct at BOP Low-Security Prisons

Aggregate Only Models, January 1998 to June 2003

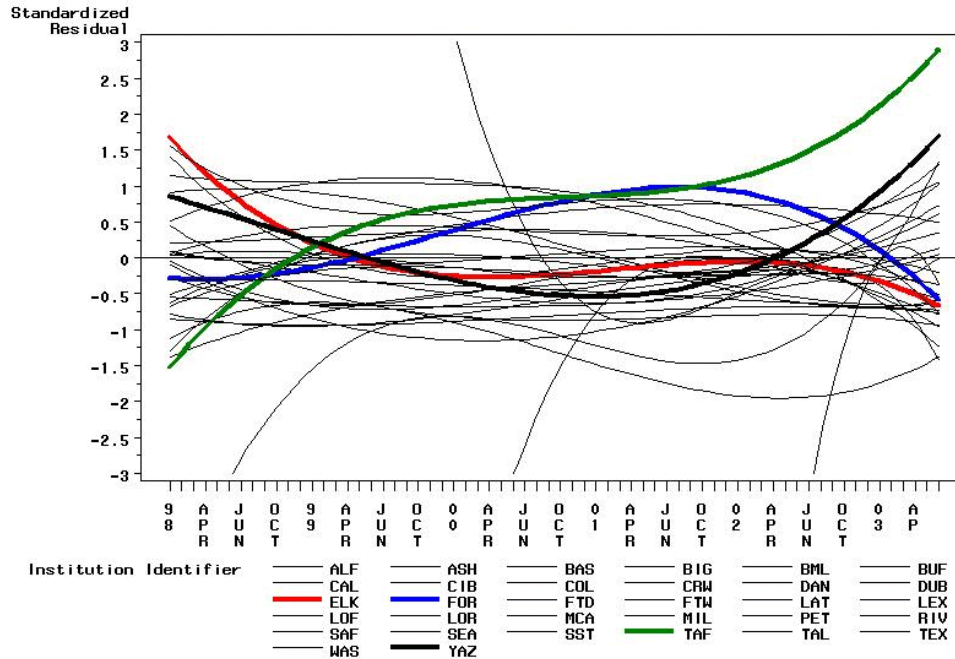
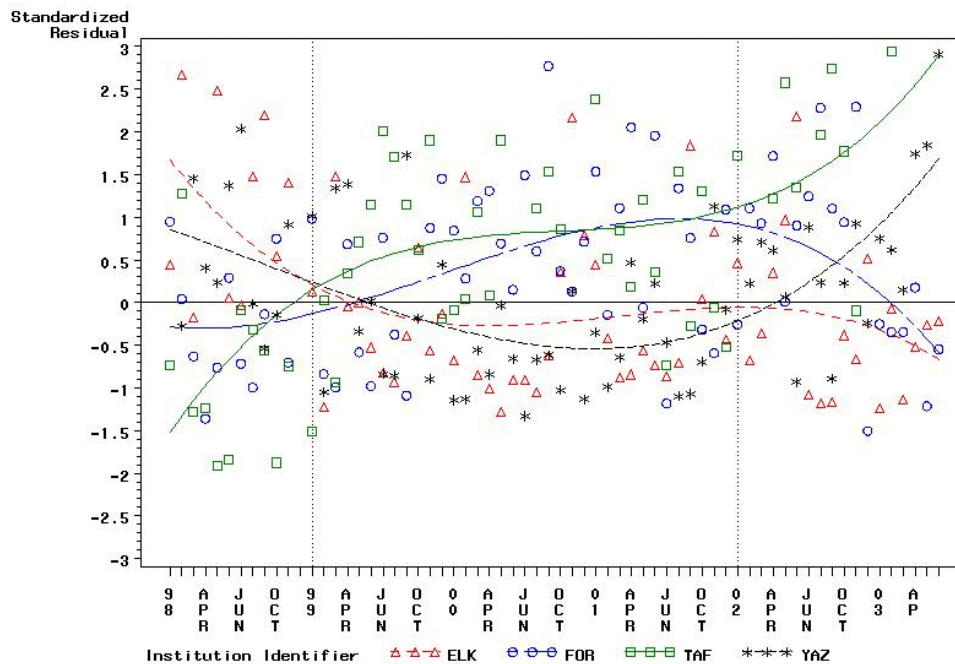


Figure IV.17

Drug Misconduct at TCI and Comparison Prisons

Aggregate Only Models, January 1998 to June 2003



Drug Misconduct at TCI, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003



200—Level Misconduct at TCI, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003



Figure IV.26

300/400—Level Misconduct at TCI and Comparison Prisons
Aggregate Only Models, January 1998 to June 2003

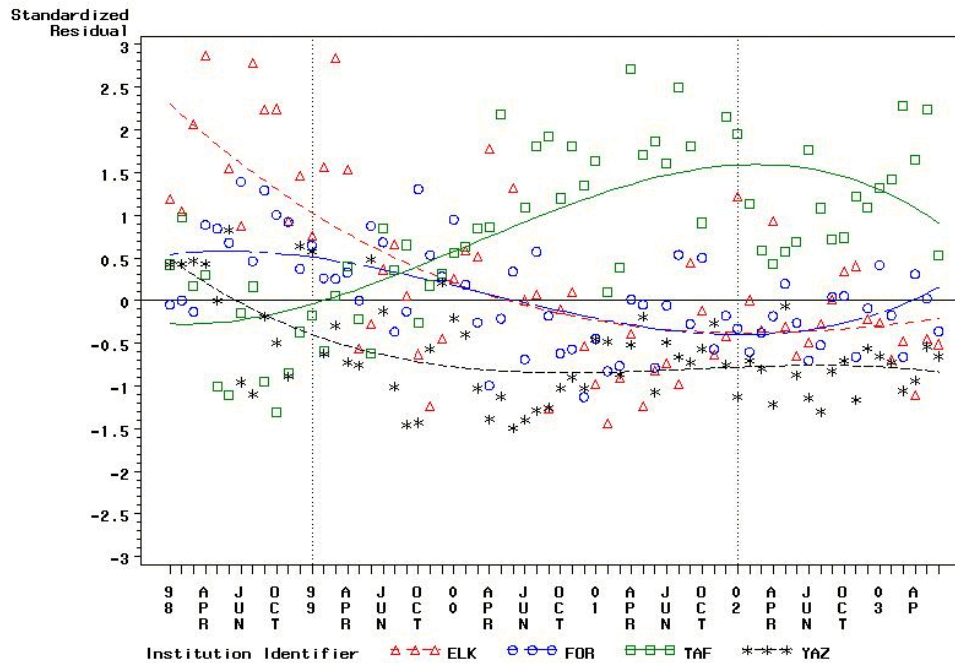


Figure IV.27

300/400—Level Misconduct at TCI, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003

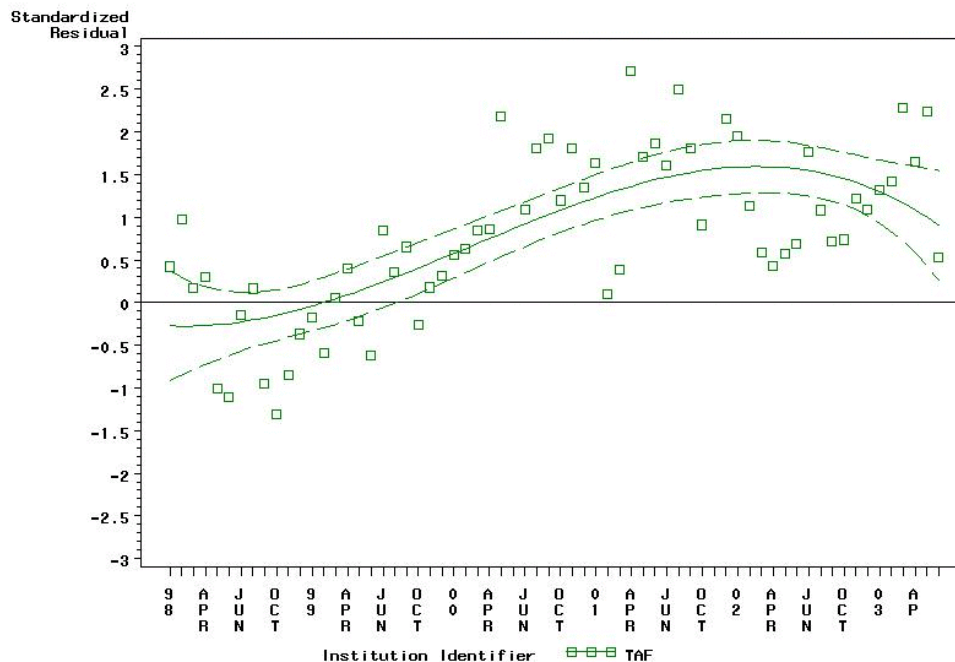


Figure IV.28

Positive Random Drug Tests Misconduct at BOP Low-Security Prisons
Aggregate Only Models, January 1998 to June 2003

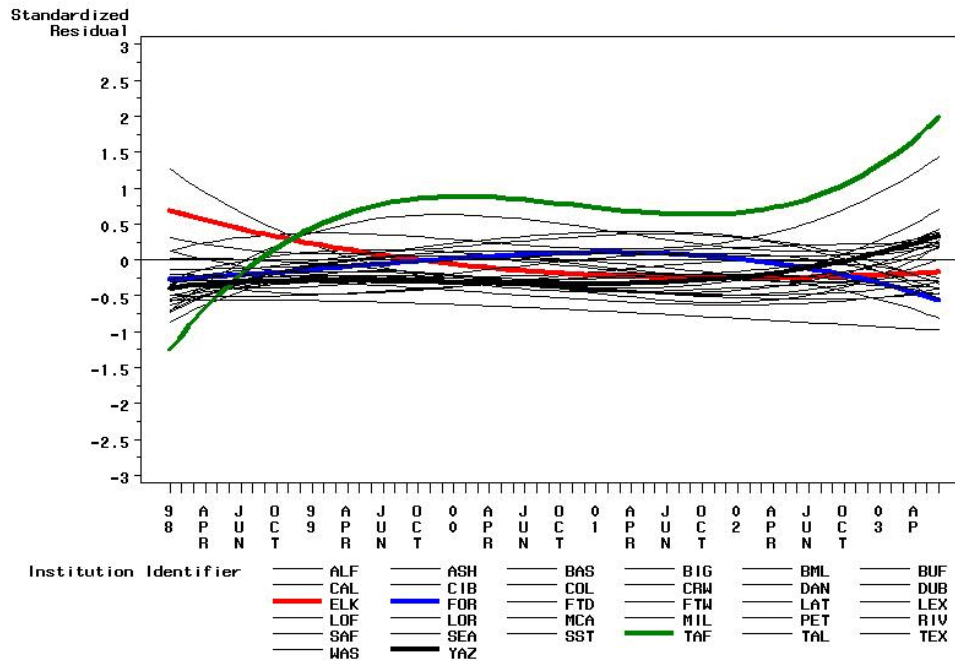


Figure IV.29

Positive Random Drug Tests at TCI and Comparison Prisons
Aggregate Only Models, January 1998 to June 2003

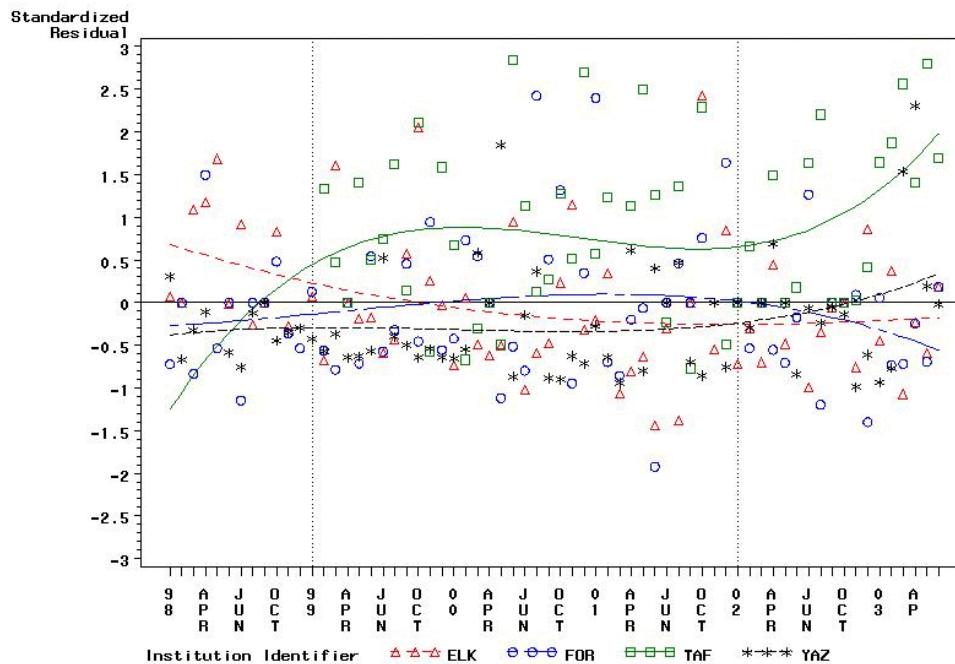
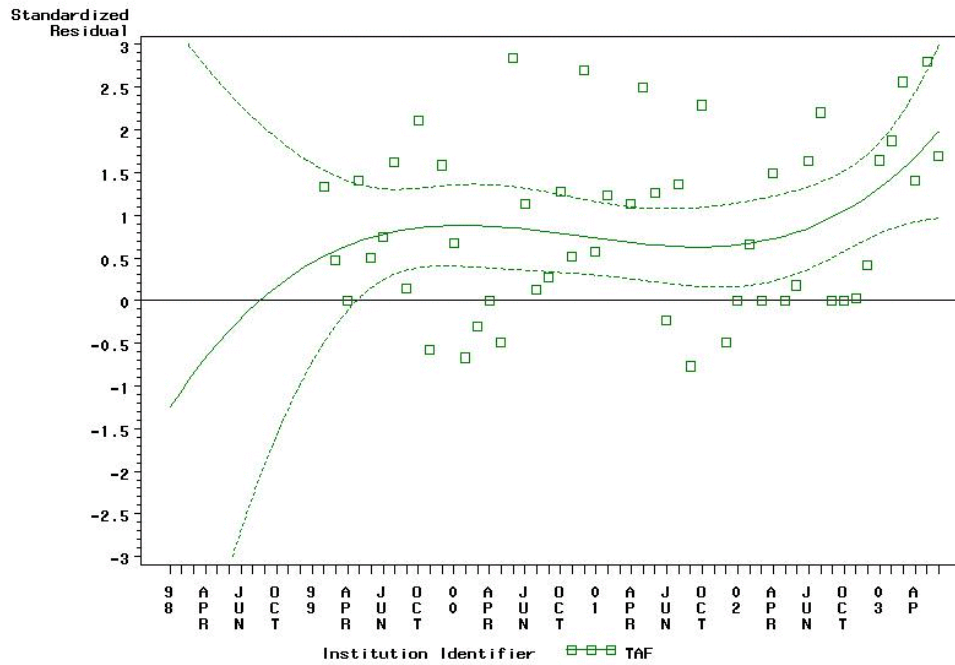


Figure IV.30

Positive Random Drug Tests at TCI, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003



SECTION V. CRITICAL INCIDENTS

This section documents incidents of inmate conduct that received special review from BOP management, either in the form of requiring an after-action report prepared by a team of correctional specialists (as in the case of a serious event such as an escape) or a daily report to Central Office correctional administrators (as with group misconduct incidents, incidents involving five or more inmates). Two events at Taft Correctional Institution (TCI) rose to the level of requiring after-action review and reporting, including an escape and a disturbance.¹⁷ TCI also experienced a second escape on December 13, 2003. The after-action report for this latter escape was actually conducted by Wackenhut and not the BOP. As will be discussed below, there were significant incidents at the BOP comparison prisons during the time period examined, although the incidents at the BOP comparison prisons did not rise to the level of requiring after-action review. TCI experienced some of these less serious events as well, as noted below.

The information in this section is presented to provide context to some of the more “dry” statistical analyses presented in earlier sections. In those sections, the details or feel for the types of behaviors occurring at the different prisons was limited. Additionally, the types of actions reported here are viewed by BOP management as being qualitatively different from individual acts of inmate misconduct. The actions reported here are the most threatening to the normal operations of the prisons with respect to insuring public safety and the safety of inmates and staff within the prisons. This section does not attempt to judge the actions of management at the different prisons, but instead the purpose is to provide a description of the actions contained in the official BOP 583 reports (reports for instances of serious misconduct) and after-action reports. For the after-action reports, the assessments provided by the after-action teams of shortcomings in institutional procedures or adherence to policy are noted.

It is important to note that the incidents recounted here only pertain to the low-security portion of the prisons. Not all of the BOP comparison prisons had an attached minimum-security prison, so the prisons were not strictly comparable on this dimension. Also, incidents requiring an after-action review are rare. The only exception is escapes. Escapes from

¹⁷Limited amounts of the information presented here on critical incidents were first reported in an article written for *Punishment & Society* (Camp, Gaes and Saylor 2002).

minimum-security prisons in the BOP are more common than escapes from secure prisons. As the preceding sentence suggests, minimum-security prisons do not have perimeter security as do all other BOP prisons. That is, there are no fences keeping the inmates on the compound, there are no gun towers, nor are their perimeter patrols by armed security officers. Both TCI and FCI Elkton had escapes from their minimum-security prisons, but these events are not discussed. Escapes from a minimum-security prison are probably more indicative of an improper classification of inmate escape risk than they are a breakdown in institutional security. Whatever factors are related to escapes from minimum-security prisons, inmates at minimum-security prisons are extremely low risks to public safety given their criminal history.

Chronology of Incidents at TCI

Incident 1. A food and work strike was held on August 25, 1999. Only 12 inmates reported for the breakfast meal, and only 36 out of approximately 270 inmates reported to their UNICOR jobs that morning (Andrews, 1999).¹⁸ The issues reported by the inmates in the interviews conducted by the contractor after the incident were the quality and quantity of food, sanitation in the kitchen, and medical concerns. This event did not require an after-action review.

TCI Incident 2. TCI experienced an inside escape from the low-security prison on September 6, 1999, a Sunday, at approximately 4:40 p.m. The inmate was able to breach two security check points in leaving the secure perimeter of the facility in the company of visitors at the conclusion of inmate visiting. On Monday, an after-action team comprised of 4 members was assembled. The after-action team included one Wackenhut employee, the Major (senior official responsible for custody) at TCI.

The team documented that the escaped inmate altered his appearance just prior to receiving a visitor on September 6. The inmate cut his hair and beard, which significantly altered his appearance from that shown on his inmate identification card. The first sign that the inmates was unaccounted for occurred at the end of visiting when the inmates identification was not picked up from the visiting room officer. After unsuccessfully paging the inmate to retrieve his identification card, the institutional staff began the process of counting and recounting the

¹⁸UNICOR is the registered trade name for prison industries in the federal government. The BOP decided to keep the prison industry operations at TCI under UNICOR control.

inmates to confirm that an inmate was missing. Following the counts and an internal search for the inmate, the Duty Warden was informed that an escape had occurred, and between 5:55 p.m. and 7:15 p.m., the Duty Warden informed the BOP, the Federal Bureau of Investigation, and local law enforcement of the escape.

The after-action team noted that during the investigation, WCC staff were extremely professional and forthcoming in providing the needed information. They did note some institutional practices that contributed to the escape. Visiting room officers were not adequately supervised or following post orders. In fact, the officers were changing roles without the knowledge of their supervisors. Visitors were not adequately supervised by escorts either coming into the institution for visiting or while leaving. One officer often escorted groups of up to 20 individuals. Visitors were escorted through the final sallyport (a pair of secure gates in a hallway that only open one at a time to prevent direct access to the outside from inside the secure area of the prison) before being identified as visitors. TCI did not have a policy to immediately reissue an inmate identification card when an inmate's appearance had been significantly altered. Proper paperwork was not being kept on the visitors or the comings and goings of the inmates. And inmate orderlies from the minimum-security prison were working in the front lobby area while visitors were being processed in and out of the prison. The orderlies dressed in a manner identical to those inmates receiving visitors, so no special concern would have been raised when the escaped inmate waited in the front lobby area prior to going outside.

TCI Incident 3. A work and food strike escalated into a general disturbance on November 15, 1999. Less lethal weapons were used in this event to bring the inmates under control, and an after-action team was assembled on November 16, 1999. This after-action team was comprised of seven BOP employees and one WCC employee, a Warden from a different WCC prison.

In an attempt to control pilferage and bring discipline to the food service operations at TCI, a new food service manager implemented changes to work conditions. The inmates responded that unless these changes were withdrawn, that they would walk off of their jobs. When the food service manager refused to rescind the changes, the inmate workers left their jobs. They were able to leave the dining hall and return to their housing units as the front doors of the dining hall were left unsecured.

At the 4:00 count, during which time inmates return to their housing units where they are locked down while the count is underway, TCI staff moved in and removed the 35 striking workers and 23 other inmates supporting the strike to the segregated housing unit (SHU). All told, 58 inmates were moved to SHU at this time.

When the institution resumed normal operations to feed the inmates the evening meal, only about 125 inmates out of approximately 1,800 actually went to the dining area. Inmates from Unit A-1A went to the dining hall and ate as scheduled, but when the next group of scheduled inmates began to move to the dining hall, inmates from Unit A-1B began pounding on the windows in their housing unit and yelling to discourage the inmates from eating. Inmates in other units also flashed the lights in their units to signal inmates not to eat in the dining hall. Even though institution staff continued releasing inmates on schedule to eat the evening meal, no inmates other than the original 125 entered the dining hall that night.

Inmates were allowed to engage in recreation that evening as well as pick up their evening medications, although the use of recreation facilities was light that evening. Instead, inmates tended to gather around the entrances to the housing units. Later in the evening, when TCI staff informed inmates that it was time to return to the housing units for the normal evening count, around 1,000 inmates milled around the outsides of the housing units and refused to enter their housing units for the 10:00 pm count. In the actions that followed, TCI staff fired several canisters of tear gas in their attempts to control the inmates. They also used sting-ball and flash bang grenades.

As early as 3:00 p.m., WCC staff realized that the possibility for an institution-wide disturbance was possible, and they had begun the process of assembling the disturbance control teams that were used to control the inmates.

The after-action team noted several shortcomings in institution procedures and practice that contributed to the disturbance. The communication with the food service workers about the change in work schedules was never formalized as should have occurred. Leaving the front doors to the dining hall open allowed the workers to leave when their demands were not met about rescinding the proposed changes. Inmates have a formal process for raising grievances with management, but it also appears that informal negotiations occurred prior to this disturbance. In fact, from the interviews, it seemed as though inmates involved in a prior food strike were not remanded to SHU. Inmates were still concerned about the quality and variety

of food, issues from the previous food strike. The inmates were released to the compound, even after it became apparent that the inmates were not going to eat the evening meal. There were other concerns as well, including bringing shotguns with lethal loads onto the compound. This error was caught by WCC managers quickly, and the unauthorized weapons were removed from the compound. Finally, local law enforcement felt that they were not kept up to date, and they wanted a debriefing of the incident.

TCI Incident 4. On October 21, 2001 two Hispanic inmates started a fight with a Black inmate over the television. After the 3 inmates were placed in SHU, multiple fights broke out between Hispanic inmates and Black inmates in the recreation yard. The Operations Lieutenant ordered the crash fences to be closed and secured. These crash fences are located in the recreation yard and separate the dormitories from the rest of the compound. Inmates jumped the fences and the fighting continued into the compound. A 911 call was placed to the CERT team and the command center was activated. The institution was placed in lock-down.

TCI Incident 5. On November 26, 2001 UNICOR work-call was announced which resulted in no one reporting for work. The institution was placed in lock-down status due to a food/work strike.

TCI Incident 6. A series of fights occurred on September 25, 2002. First, a number of fights broke out in the recreation yard, as staff were responding to that incident, a number of fights started in front of the Education building and then moved to outside of the religious services and food services areas. All inmates were sent back to their units, and the institution was locked down. A preliminary investigation revealed that two white inmates had assaulted a Hispanic inmate early in the day, possibly due to drugs. Later in the evening, Hispanic inmates retaliated and assaulted white inmates. As a result of these fights 15 inmates were locked up, and they anticipated locking up an additional 21 inmates. It was reported at 11:30 pm a total of 18 inmates were locked up.

TCI Incident 7. An escape from the secure, low-security portion of TCI occurred on December 13, 2003. A Mexican national inmate serving a 210 month sentence (March 2011 release date) for the distribution of methamphetamine hid himself in a trash compactor behind the food services rear dock. The trash was picked up at about 1:00 pm by a local contractor who dumped the trash at a landfill about 10 miles from the facility. After dumping the trash,

the contractor noticed a person with gray sweats running through the landfill and immediately contacted TCI staff. Wackenhut staff immediately locked down the facility, performed an inmate count, and found that an inmate was missing. They dispatched a search team to the town near the landfill, and they saw the inmate enter a food market. They approached the inmate who surrendered without incident, and the Wackenhut search team returned the inmate to the facility where he was placed in the special housing unit.

Chronology of Events at BOP Comparison Prisons

The BOP comparison institutions have not been free of critical incidents during this time period, although there have been no incidents that required the use of an after-action review team. Many of the critical incidents at BOP comparison prisons occurred at one facility, FCI Forrest City, where there was a problem with Hispanic gangs. While gangs have been a factor for the BOP for many years, only recently have gangs been active in lower-security prisons. In the past, gang activity was largely confined to higher-security prisons, especially penitentiaries. The BOP experienced problems with Hispanic gangs at other BOP low-security prisons during 1999 as well, especially in the South Central Region. As shown below, the other BOP comparison institutions had their own issues during this time period.

FCI Forrest City

Incident 1. On January 24, 1998, there was a fight at FCI Forrest City involving six Hispanic inmates. An ink pen and a combo lock were used as weapons. Staff separated the inmates involved and escorted them to the Lieutenant's Office for further questioning. Inmates were then escorted to Medical Services for treatment of their injuries and placed in the special housing unit (SHU).

Incident 2. On January 26, 1998, there was a work and food strike. The work strike lasted for one day, but the food strike went on for two days. Meetings and interviews were conducted with numerous inmates. On January 27, 2003, a recall was announced and inmates reported back to their housing units for an institutional lock down. Approximately 140 inmates attended the noon meal. During the evening meal, approximately 1,160 inmates participated in the action. On January 28, 2003, normal operations resumed.

Incident 3. On February 10, 1998, a body alarm was activated for assistance. Three Black inmates stated that they were assaulted by a group of five Hispanic inmates. The inmates were escorted to Health Services for evaluation and treatment. The three inmates were placed in administrative detention pending further investigation.

Incident 4. On March 21, 1998, an inmate reported he had been assaulted by unidentified inmates. An investigation into the assault revealed 4 inmates as possible assailants. All inmates were escorted to Health Services for medical evaluation. All identified inmates were placed in SHU pending further investigation.

Incident 5. On March 21, 1998, a body alarm was activated due to a physical altercation involving approximately nineteen (19) inmates which was occurring in and around the Multi-Purpose Room. Responding staff observed several unidentified inmates exiting the upper floor and moving across the compound toward the other two housing units. Three inmates were observed with obvious injuries, and they were removed and escorted to medical services for evaluation and treatment. All inmates were examined for signs of injury during the 4:00 pm count, and those inmates displaying injury were removed and seen by medical staff. All identified inmates were subsequently placed in the SHU.

Incident 6. On March 1, 1999, there was a disturbance between thirty-four Hispanic inmates and fifteen Black inmates. The Unit Officer in one of the housing units noticed that inmates had grouped within the unit and were arguing. When responding staff arrived on the scene, they observed one of the Black inmates swinging a plastic chair at the Hispanic inmates. When he refused an order to stop, the inmate was restrained by institution staff. An institution shutdown was instituted since many housing areas are open and a lockdown was not possible. Two staff and ten inmates received minor injuries, and forty-two inmates were moved to the SHU.

Incident 7. On April 6, 1999, there was a fight between 40 to 50 Hispanic inmates in the recreation yard. Some of the inmates were armed. Staff were able to initially identify 17-19 of the inmates involved. These inmates were restrained and escorted to the Health Services Unit where they were interviewed and examined by medical staff. The inmates were all gang members belonging to rival gangs, the Border Brothers and Barrio Aztecas. Twenty-four inmates were moved to the SHU.

Incident 8. On April 21, 1999, information was received that a possible work strike was to take place on May 3, 1999. An investigation was conducted and concluded on April 25, 1999, and a total of eight inmates were placed in SHU. On May 3, 1999, there was a work strike at FCI Forrest City. Only 8 inmates reported to work following the morning meal. An institution recall was announced, and a count was conducted. Inmates were confined to their units with restricted movement. Mass inmate interviews were conducted throughout the morning. By 1:00 that day, when a general work call was announced, all inmates, with the exception of ten, reported back to work.

Incident 9. On August 1, 1999, there was a fight between seven Hispanic inmates and five Jamaican inmates in the patio area of the recreation yard. Staff did not observe the fight, but blood was noticed in the recreation area. Four of the inmates were treated at an outside medical facility for the injuries sustained during the fight. One inmate lost consciousness during the interview process. A total of twelve inmates were admitted into the SHU for their participation in this action.

Incident 10. On August 26, 1999, a body alarm was activated after a fight involving two inmates in the rear of the housing unit. Responding staff found one inmate secured in the Officer's Station and escorted him to the Health Services Unit for evaluation. The second inmate was found locked in a shower stall. He was escorted from the Unit to Health Services for evaluation. Both inmates were subsequently transported to an outside medical facility for treatment. Approximately 2 hours later, the Unit Officer again activated his body alarm. An inmate was found in the shower bleeding. He stated that he had been attacked by several unidentified inmates. He was transported to an outside medical facility for treatment. A preliminary investigation reveals the two incidents are related and an investigation continues.

Incident 11. On November 10, 2000, there was a fight between one black inmate and five Hispanic inmates. The fight took place in a recreation cage.

Incident 12. On July 12, 2001, three members of the Hispanic gang, Mexicanemi, were assaulted by ten other inmates.

Incident 13. On July 13, 2001, four inmates assaulted a single other inmate.

Incident 14. On July 20, 2003, one inmate was assaulted by 4 other inmates with homemade sharpened weapons and a mop wringer. The inmate sustained five superficial puncture wounds to the back, to the top of his hand, and one through his middle finger. He was escorted to the local hospital for further treatment. All inmates were placed in Administrative Detention.

FCI Elkton

Incident 1. There was a food strike that lasted for two days that started on October 27, 1998. The precipitating issue was food portions.

Incident 2. On July 7, 1998, nineteen inmates who were in line at the commissary refused to report to their assigned areas at the close of the open movement period. Responding staff arrived and escorted 19 inmates to SHU without further incident. The action was classified as a demonstration.

Incident 3. On October 19, 1998, the Operations Lieutenant began calling the rotation for the feeding of the noon meal. The compound was cleared, and UNICOR inmates were called to the dining hall. Twenty inmates from UNICOR reported to the dining hall for the meal. The Operations Lieutenant then notified the Captain of the response from UNICOR and continued with the meal rotation. Only eleven inmates reported to the dining hall from the housing units. The inmate population was then recalled to the housing units in preparation for a count and mass interviews. The emergency count was initiated, and the count cleared. Eleven inmates were placed in SHU due to their disruptive behavior during count. Mass interviews were conducted and inmates reported issues such as visiting complaints and rumors of a pending population increase at FCI Elkton as the primary reason for the food boycott.

Incident 4. On March 18, 1999 staff were responding to a body alarm in the secretary's office. The staff attempting to exit a unit to respond to the body alarm ordered the inmates in a stairwell to move out of their way. The inmates stepped out onto the compound and began yelling and jeering at staff responding to the body alarm. After the body alarm was cleared, the Lieutenant was speaking to the inmates involved. Another inmate approached the front of the unit wielding a broom and a long handle dustpan, cursing at staff, and speaking to the other inmates in Spanish. The inmate raised the broom in the air in an attempt to strike a staff member. The inmate was grabbed by several staff members, and even though the broom and

dust pan were removed, the inmate continued to struggle. When the Operations Lieutenant approached the inmate, he began shouting obscenities and attempted to hit the Operations Lieutenant with a closed fist. The Lieutenant grabbed the inmate's hand and attempted to apply restraints. The inmate continued to struggle, and staff placed the inmate on the ground in an attempt to apply restraints and keep control of the inmate. The Lieutenant called for assistance. Meanwhile, the other 5 inmates present began to move toward the area. Another inmate had to be physically restrained to keep him from assaulting the Lieutenant. All six inmates were escorted to SHU without further incident.

Incident 5. On October 15, 1999 a white inmate had an argument with an unidentified Hispanic inmate for cutting in front of him in line during the evening meal. Later in the evening, the white inmate was assaulted by 6 to 7 unidentified Hispanic inmates near the inmate phones. The inmate was escorted to Health Services for a medical examination. He received a laceration on top of his head and swelling to his right hand.

Incident 6. On February 12, 2000 a body alarm was activated, and six inmates were observed fighting each other with a large crowd of inmates watching. Staff entered the housing unit and ordered the inmates to stop fighting. Four of the inmates ceased and were restrained. The other two inmates continued striking each other in the face with closed fists. All inmates were escorted to the institution hospital for examinations and treatment. During the medical examination, staff escorted one inmate in the vicinity of another inmate, and the second inmate attempted to kick the inmate who was being escorted. All inmates involved in this incident were under the influence of intoxicants. During a search of the unit, an additional six inmates, were observed to be under the influence of intoxicants, and they were placed in SHU after being examined by medical staff.

Incident 7. On March 18, 2000, two Hispanic inmates were fighting under the smoke stack. Upon staff arrival, one inmate was able to flee the area by blending with the crowd of inmates who had gathered in the area at the time of the fight. The inmate who fled left his jacket lying on the ground with his identification in it. As staff were searching for the inmate, another call came in stating that 10 Hispanic inmates were fighting. Staff ordered the inmates to stop fighting and lay on the ground. All inmates complied and were escorted to the institution Health Services for examination. All of the inmates received only minor abrasions, and they were then placed in SHU pending further investigation.

Incident 8. On July 21, 2000, around 40 inmates (approx. 20-30 African-American inmates and 10 Mexican inmates) were involved in a fight in the recreation yard. When staff responded, they found one inmate unresponsive. This inmates was later found to have a fractured skull. Another inmate had sustained a fractured eye socket. Once all areas had been cleared, staff members began to conduct a mass shakedown of the recreation area. During the search, the following items were found: a shank (homemade knife), a cane, and 3 lock-in-sock weapons. Twenty inmates were placed in SHU pending further investigation.

Incident 9. On October 25, 2000, one inmate was assaulted by six inmates. All inmates were placed in SHU pending further investigation.

Incident 10. On October 20, 2001, a single inmate was assaulted by five African American inmates. The five inmates conducting the assault were placed in SHU.

Incident 11. On November 12, 2001, a single African American inmate was assaulted by five African American inmates. All inmates were sent to SHU.

Incident 12. On June 22, 2002, the Operations Lieutenant was contacted by a correctional officer and asked to come to the unit. Upon his arrival, the Lieutenant noticed blood spots on the floor in front of the multi-purpose room. The unit was locked down, and a search of the unit was conducted by staff. An inmate was found in the shower with multiple injuries to his head and upper body. He was treated by the Duty PA for multiple superficial puncture wounds and abrasions. Later, he was transported to an outside hospital for a follow-up examination. A preliminary investigation revealed that the inmate was assaulted with a homemade sharpened weapon and fists. Five Hispanic inmates were placed in SHU pending further investigation.

FCI Yazoo City

Incident 1. On June 3, 2000, the institution was placed on lockdown (all inmates confined to their housing units) because of three fights between Hispanic and African American inmates. The seven Hispanic inmates involved were all members of Sureños, a Hispanic gang. The African Americans were from Alabama and Mississippi. The fights were caused by the belief that the five African American inmates involved had stolen pornography from the Hispanic inmates.

Incident 2. On August 8, 2000, ten Jamaican inmates assaulted another inmate in the recreation yard. The fight was broken up by staff, with no life threatening injuries. Intelligence gathering revealed the group of 10 Jamaicans jumped on the inmate because they believed he was a snitch. All inmates were taken to SHU pending further investigation.

Summary

The critical incidents reviewed here suggest that there may be differences between the four prisons with respect to the types of misconduct that rise to the next level of scrutiny. In particular, it appears that institutional executive staff at TCI faced serious issues related to institutional operations and safety. Escapes from a secure prison are a very serious matter with respect to public safety, and TCI experienced two escapes within a short period of time. During the same time period, the BOP operated well over 100 different prison locations, many with multiple prisons per site, and only experienced three escapes from secure prisons involving five inmates.¹⁹ One of the escapes was from the converted prison camp at FCI Elkton in 2001. Initially, the housing unit outside of the secure perimeters of the low-security prison was operated as a minimum-security prison camp. However, it was converted to a Federal Satellite Low (FSL) facility in 2001. FSL facilities hold inmates with minimum-security classifications, but who have public safety issues that prohibit them from being placed in minimum-security prisons. While more secure than minimum-security prisons, for example, the perimeter fencing is much like at a low-security prison, FSL prisons do not have all of the security measures of low-security prisons. The other two escapes from BOP prisons occurred at medium-security prisons. Likewise, despite the fact that comparison prisons experienced food and work strikes, none of the food or work strikes at comparison prisons escalated into a general disturbance as happened at TCI. The coincidence of three very serious events at one prison over the course of six years raises some level of concern.

¹⁹Since the first draft of this report was written, the BOP experienced two more escapes in 2004, one from a low-security prison and one from a federal medical center.

SECTION VI. CONCLUDING REMARKS

This report was designed to address the experiences of the BOP with the private prison in Taft, California that is operated by Wackenhut Corrections Corporation. The report reached conclusions about the performance of Taft Correctional Institution. It is important to note that cost and performance issues should be seen as opposite sides of the same coin and jointly examined. While cost was not examined in this report, it was examined by Nelson (2005) for these same facilities and for the same time period. She found that government operations of TCI would have cost about the same as operations by Wackenhut; about 2.6 percent higher under government control. This is important for the quality evaluation provided here in that the government was not spending more at the comparison prisons in ways that would have provided an unfair advantage for the quality comparisons.

While TCI generally meets the terms of the contract with the BOP, several sources of information suggest that performance at TCI was problematic in a relative sense, e.g., that performance did not compare favorably to similar BOP prisons. No attempt was made in this report to establish baseline levels of acceptable performance, with the exception of the section on contract issues and contract compliance. For the most part, TCI performed at an acceptable level regarding the terms of the contract, and such performance improved over the initial five years of the contract with respect to TCI receiving fewer contract deductions and more of the award fees. Baseline levels, though, were not set for inmate misconduct and drug use; such attempts tend to be subjective and dependent upon the values of those setting the baseline. Instead, all performance comparisons were relative to the performance of other prisons. It is entirely possible that the performance of all prisons examined in this report met the baseline standards of correctional practice. In fact, since BOP prisons and TCI are accredited by the American Correctional Association (ACA), it is likely that all prisons indeed did meet the baseline standards for correctional practices codified by ACA.

Having noted that the prisons probably met ACA standards, the fact remains that several trends at TCI were problematic. First, the review of significant incidents at TCI revealed that TCI experienced three events of critical importance. There were two escapes from the secure main facility, and there was a disturbance during which around 1,000 inmates refused to return to their housing units for the 10:00 p.m. count. The escapes threatened public safety, and the disturbance had the potential to threaten public safety and surely threatened the safety

of inmates and staff at TCI. None of the comparison prisons had comparable events, and there was only one escape from any BOP prison during the five-year period examined here. During this time, the BOP operated well over 100 different prison sites, and some of the sites contained more than one prison.

Second, the counts of all forms of misconduct considered together were higher at TCI than was expected given the characteristics of the inmate population at TCI. In fact, for all misconduct, TCI had some of the largest deviations from the expected counts, suggesting that performance at TCI was problematic. When all misconduct was broken into three categories based upon severity of the behavior, it was seen that the trend for overall counts of misconduct was also observed for the most serious category of misconduct, 100-level offenses, and the least serious category, the 300- and 400-level offenses. The trend was along expected lines for 200-level categories of misconduct.

Third, the results of random drug tests at TCI also suggested that performance was a problem. The counts of positive drug tests were the highest of the BOP low-security prisons for much of the period examined. Data on random drug tests are important because they measure actual prevalence of the behavior and reflect upon the ability of the institution to implement procedures and practices to limit the behavior.

All of the performance data considered simultaneously—the critical incidents experienced at TCI, the higher than expected counts of misconduct, and the higher than expected counts of positive results for random drug tests—suggest that performance at TCI was more problematic than performance at the BOP comparison prisons and other BOP low-security prisons. With the exception of the critical incidents, it is not clear that the performance was qualitatively different from that of BOP low-security prisons in general. It would be more defensible to claim that TCI performed at a level consistent with other low-performing BOP prisons. The two escapes, and the disturbance, though, were exceptional in comparison to BOP prisons.

The findings of low performance of TCI in comparison to BOP prisons were strengthened by the fact that costs of operations at TCI and the BOP comparison prisons were generally comparable (Nelson 2005). The other BOP low-security prisons spent more per inmate day than did TCI and the BOP comparison prisons, and this additional expenditure of resources potentially confounded the performance results presented in Section V. When comparing TCI to the BOP comparison prisons, though, there is no confounding effect of the BOP spending

more resources to “purchase” greater quality. In fact, TCI used considerably more staff than the BOP comparison prisons but was unable to perform at the higher levels generally noted at the BOP comparison prisons.

It could be argued that the additional staff resources at TCI provided the conditions for greater discovery of misconduct at TCI than at the BOP comparison prisons. However, there is an alternative point of view about additional staff resources. Most correctional professionals maintain that additional staff resources deter misconduct because additional staff provides greater surveillance. Unfortunately, monthly staff data were not available for TCI, so it was not possible to empirically investigate the effect of inmate-to-staff ratios on inmate misconduct in this report. For violent misconduct, though, preliminary analyses at the BOP found that higher staff-to-inmate ratios were associated with lower rates of misconduct, not higher rates (internal memo written by William G. Saylor on February 2004).

It is likely, then, that the higher rates of inmate misconduct at TCI were not a statistical artifact of differences in inmate-to-staff ratios. This assumption is buttressed by the high hit rates for random drug tests at TCI, the significant incidents recorded there (escapes and a disturbance), and the higher rates of 100-level misconduct. This leaves the overall conclusion that TCI cost about the same as similarly-sized BOP prisons, but the quality of operations at TCI was more of an issue, even though TCI generally met the conditions of the contract including ACA accreditation.

APPENDIX 1. MISCONDUCT CODES AND CATEGORIZATIONS

APPENDIX 1. MISCONDUCT CODES AND CATEGORIZATIONS

Code	Short Description	Category 1**	Category 2
100	Killing	Violence	Violence
101	Assaulting with Serious Injury	Violence	Violence
102	Escaping—Secure Custody or with Violence	Security	
103	Setting a Fire	Violence	
104	Possessing A Dangerous Weapon	Violence	Violence
105	Rioting	Violence	Violence
106	Encouraging Others to Riot	Violence	
107	Taking a Hostage	Violence	Violence
108	Possessing a Hazardous Tool	Violence	
109	Possessing Drugs/Drug Items (Replaced with 113)	Drug	Drug
110	Refusing to Take a Drug Test	Drug	Drug
111	Introduction of Drugs/Drug Items	Drug	Drug
112	Use of Drugs/Drug Items	Drug	Drug
113	Possessing Drugs/Drug Items	Drug	Drug
197	Phone Abuse, Criminal	Security	
198	Interfering with Staff—Greatest Severity	Security	
199	Disrupting Conduct—Greatest Severity	Security	
200	Escaping	Accountability	
201	Fighting with Another Person	Violence	Violence
203	Threatening Bodily Harm	Violence	Violence
204	Extorting/Blackmail/Protecting	Violence	
205	Engaging in Sexual Acts	Sexual*	
206	Making Sexual Proposal/Threat	Sexual*	
207	Wearing a Disguise or Mask	Security	
208	Interfering with Security Devices	Security	
209	Adultering Food or Drink	Property	
211	Possessing Staff Clothing	Security	
212	Engaging in Group Demonstration	Security	

APPENDIX 1. CONTINUED

Code	Short Description	Category 1	Category 2
213	Encouraging Refusal of Work	Security	
215	Introducing Alcohol into Facility	Drug	
216	Bribing Official, Staff Member	Security	
217	Exchanging Money for Contraband	Security	
218	Destroying Property over \$100	Property	
219	Stealing	Property	
220	Using Martial Arts/Boxing	Violence	
221	Being in Unauthorized Area with Opposite Sex	Accountability	
222	Possessing Intoxicants	Drug	Drug
223	Refusing to Take Alcohol Test	Drug	Drug
224	Assaulting without Serious Injury	Violence	Violence
297	Phone Abuse, Non-Criminal	Security	
298	Interfering with Staff–High Severity	Security	
299	Disruptive Conduct–High Severity	Security	
300	Indecent Exposure	Sexual*	
302	Misusing Medication	Drug	Drug
303	Possessing Unauthorized Money	Property	
304	Lending for Profit	Property	
305	Possessing Unauthorized Item	Property	
306	Refusing Work/Program Assignment	Accountability	
307	Refusing to Obey an Order	Accountability	
308	Violating a Condition of Furlough	Other	
309	Violating a Condition of a Community Program	Other	
310	Being Absent from Assignment	Accountability	
311	Failing to Work as Instructed	Accountability	
312	Being Insolent to Staff Member	Other	
313	Lying or Falsifying Statement	Other	
314	Counterfeiting or Forging Document	Property	

APPENDIX 1. CONTINUED

Code	Short Description	Category 1	Category 2
315	Participating in Unauthorized Meeting	Security	
316	Being in Unauthorized Area	Accountability	
317	Failing to Follow Safety Regulations	Accountability	
318	Using Unauthorized Equipment/Machinery	Security	
319	Using Equipment Contrary to Instructions	Security	
320	Failing to Stand Count	Accountability	
321	Interfering with Taking Count	Security	
324	Gambling	Property	
325	Conducting a Gambling Pool	Property	
326	Possessing Gambling Paraphernalia	Property	
327	Contacting Public without Authorization	Security	
328	Giving/Accepting Money without Authorization	Property	
329	Destroy Property \$100 or Less	Property	
330	Being Unsanitary or Untidy	Other	
331	Possessing a Non-Hazardous Tool	Property	
332	Smoking in Unauthorized Area	Accountability	
397	Phone Abuse, Non-Criminal	Security	
398	Interfering with Staff—Moderate Severity	Security	
399	Disruptive Conduct—Moderate Severity	Security	
400	Possessing Unauthorized Property	Property	
401	Possessing Unauthorized Amount of Clothing	Property	
402	Malingering, Feigning Illness	Accountability	
403	Smoking in Unauthorized Area	Accountability	
404	Using Abusive/Obscene Language	Accountability	
405	Tattooing or Self-Mutilation	Other	
406	Using Phone or Mail without Authorization	Security	
407	Violating Visiting Regulations	Security	
408	Conducting a Business without Authorization	Security	

APPENDIX 1. CONTINUED

Code	Short Description	Category 1	Category 2
409	Unauthorized Physical Contact	Sexual*	
497	Phone Abuse, Non-Criminal	Security	
498	Interfering with Staff—Low to Moderate Severity	Security	
499	Disruptive Conduct—Low to Moderate Severity	Security	

* Two categorizations of offenses were used in the analyses in Section V. The first categorization was used for the analyses of individual-level data where the importance of individual-level variables was demonstrated for the multilevel models. It is a complete categorization of BOP offenses. The second categorization was used to compare the multilevel and aggregate-only derivations of institutional performance. In the Key Indicators/Strategic Support System from which the aggregate data were drawn, many BOP offense codes are collapsed into “other” categories. The second categorization is based upon the same logic as the first categorization, but offense codes that were not separately identifiable in the KI/SSS data system were dropped from the respective categories. Also, only the offense categories of drug-related misconduct and violent misconduct were categorized in this latter fashion.

** Sexual misconduct was not analyzed in the current report. Sexual misconduct simply did not have a multilevel structure.

APPENDIX 2. ADDITIONAL MULTILEVEL PERFORMANCE MEASURES

Figure 2.1

Accountability Misconduct at BOP Low-Security Prisons
Multilevel Models, January 1999 to December 2001

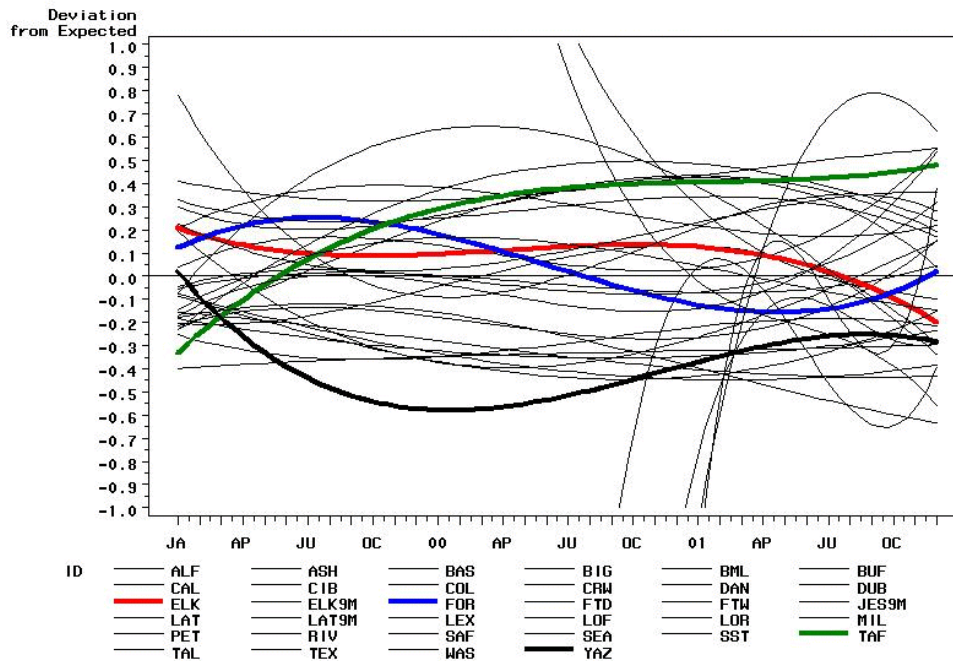


Figure 2.2

Accountability Misconduct at TCI and Comparison Prisons
Multilevel Models, January 1999 to December 2001

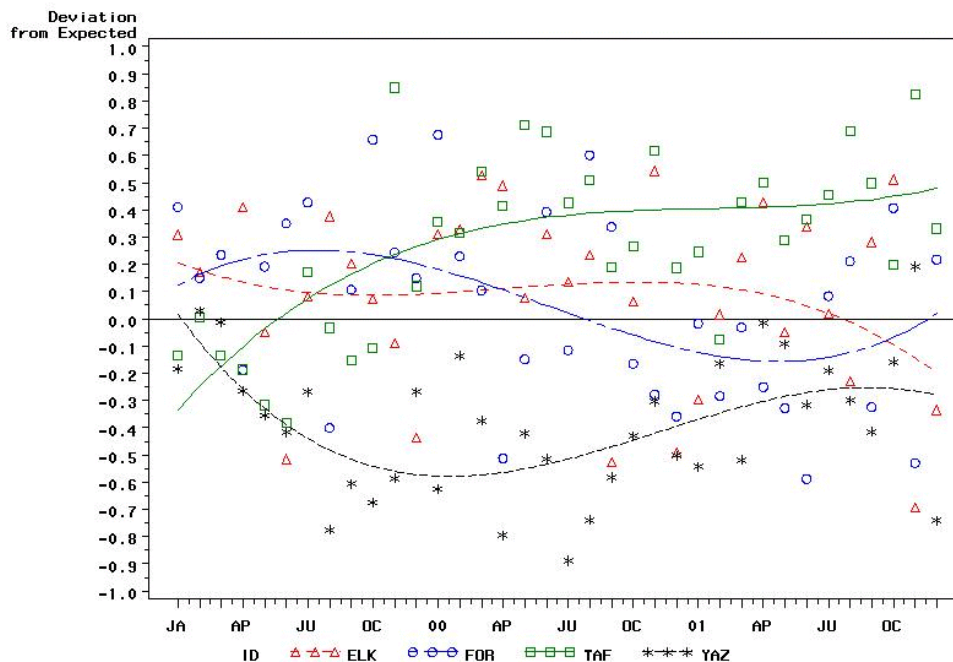


Figure 2.3

Accountability Misconduct at TCI, with 95% Confidence Interval
Multilevel Models, January 1999 to December 2001

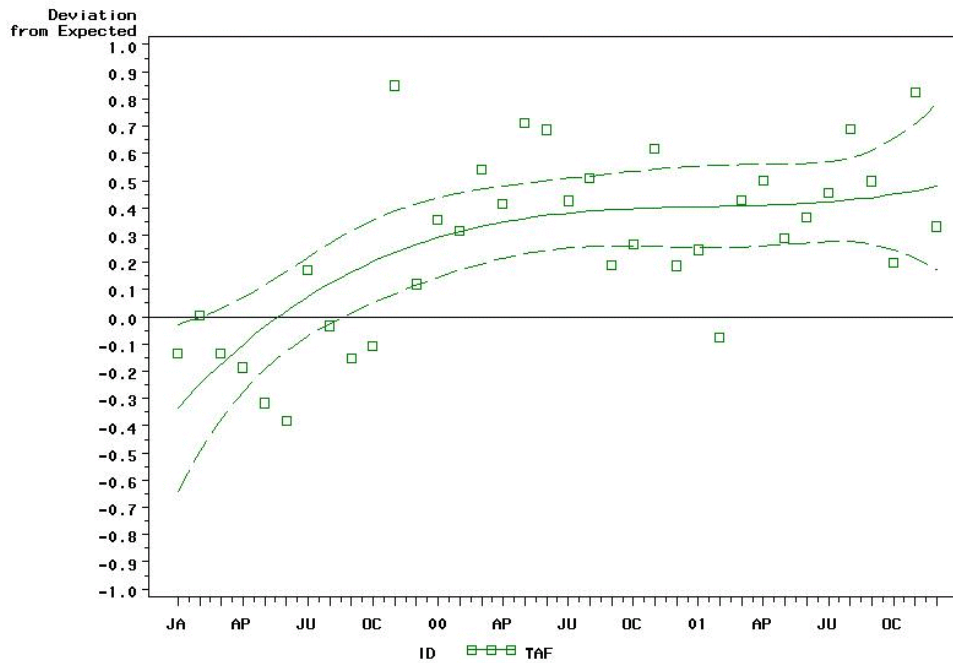


Figure 2.4

Security Misconduct at BOP Low-Security Prisons
Multilevel Models, January 1999 to December 2001

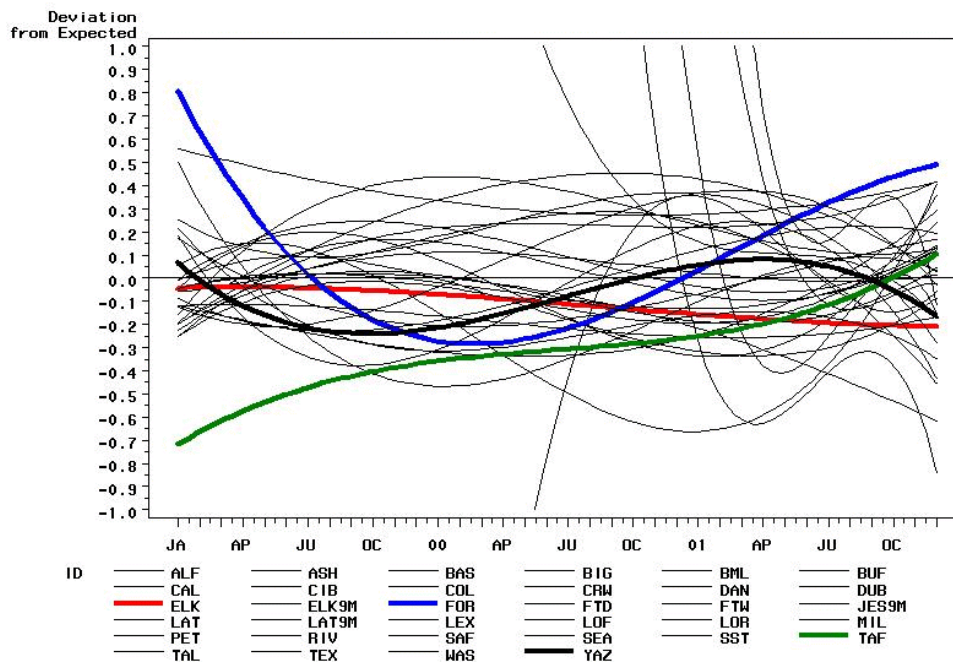


Figure 2.5

Security Misconduct at TCI and Comparison Prisons
Multilevel Models, January 1999 to December 2001

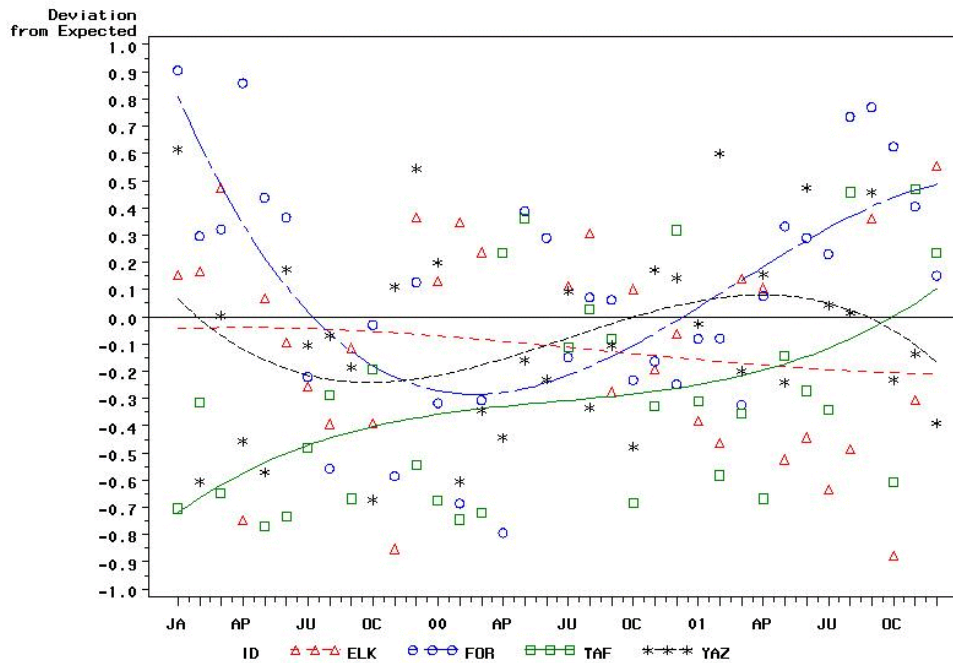


Figure 2.6

Security Misconduct at TCI, with 95% Confidence Interval
Multilevel Models, January 1999 to December 2001

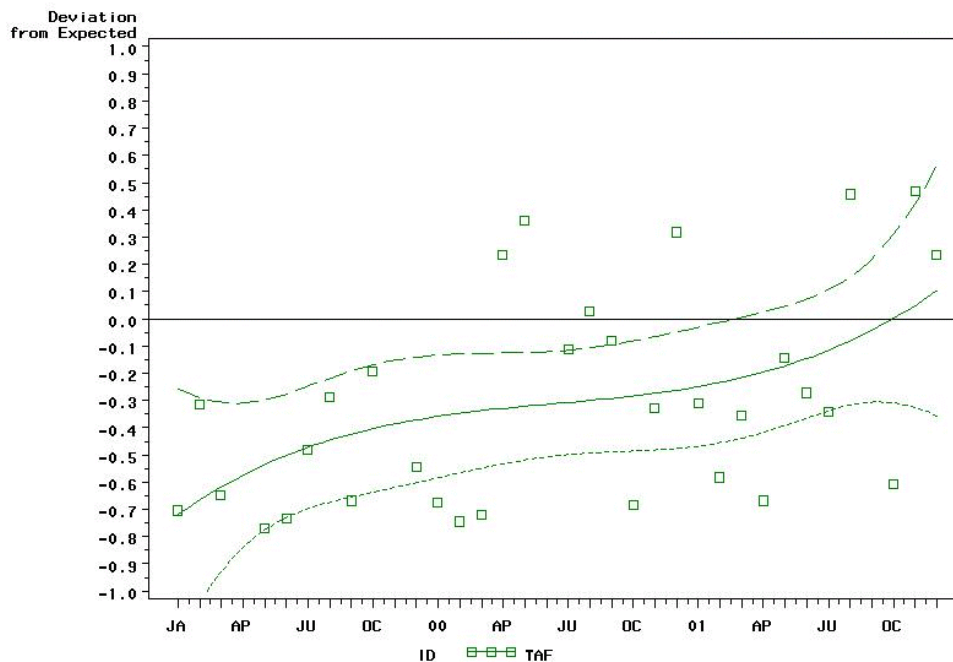


Figure 2.7

Property Misconduct at BOP Low-Security Prisons
Multilevel Models, January 1999 to December 2001

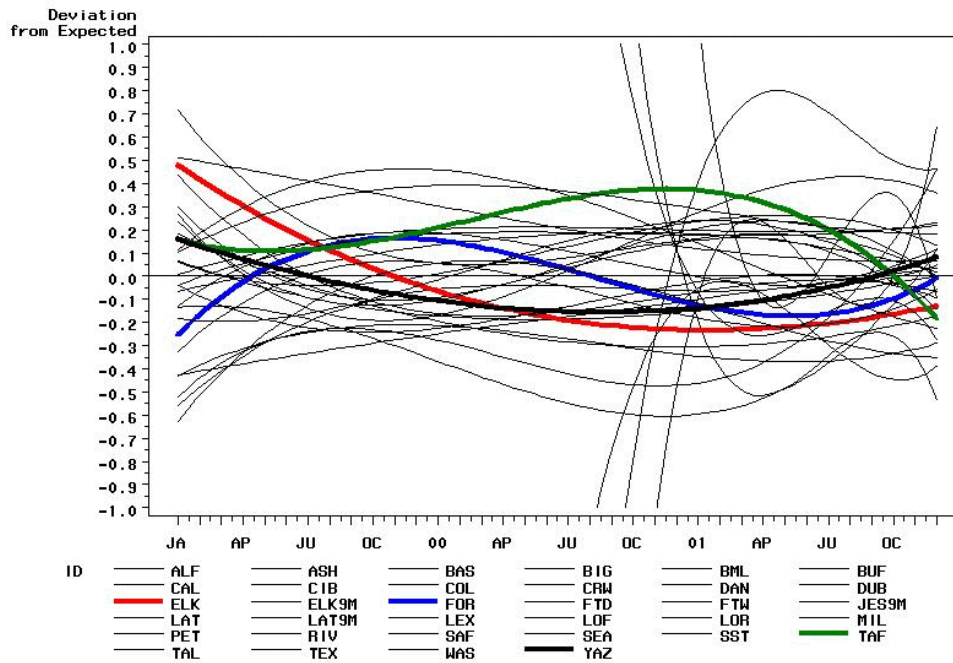


Figure 2.8

Property Misconduct at TCI and Comparison Prisons
Multilevel Models, January 1999 to December 2001

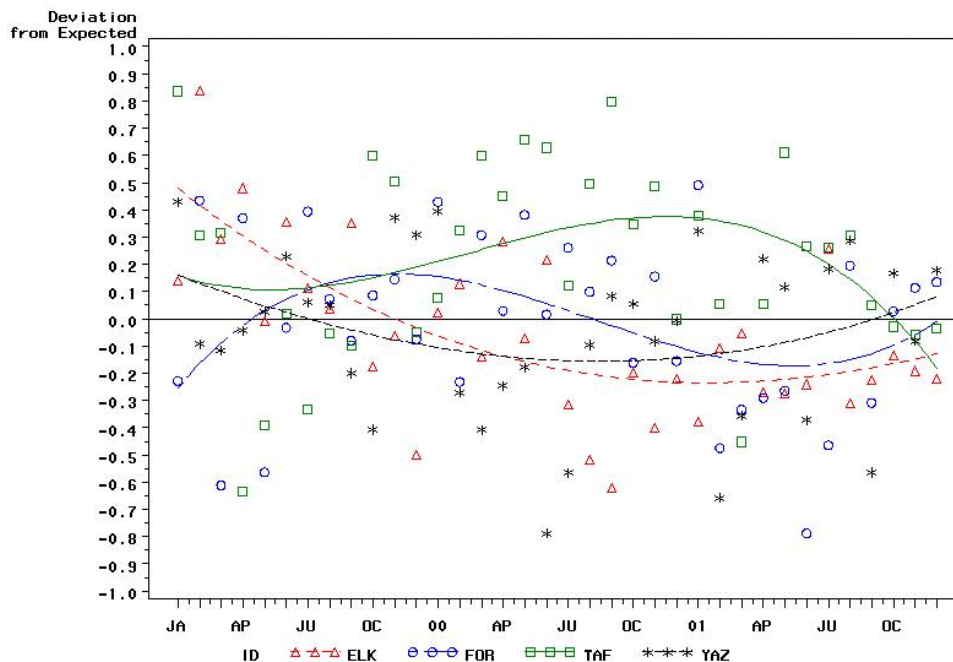


Figure 2.9

Property Misconduct at TCI, with 95% Confidence Interval
Multilevel Models, January 1999 to December 2001

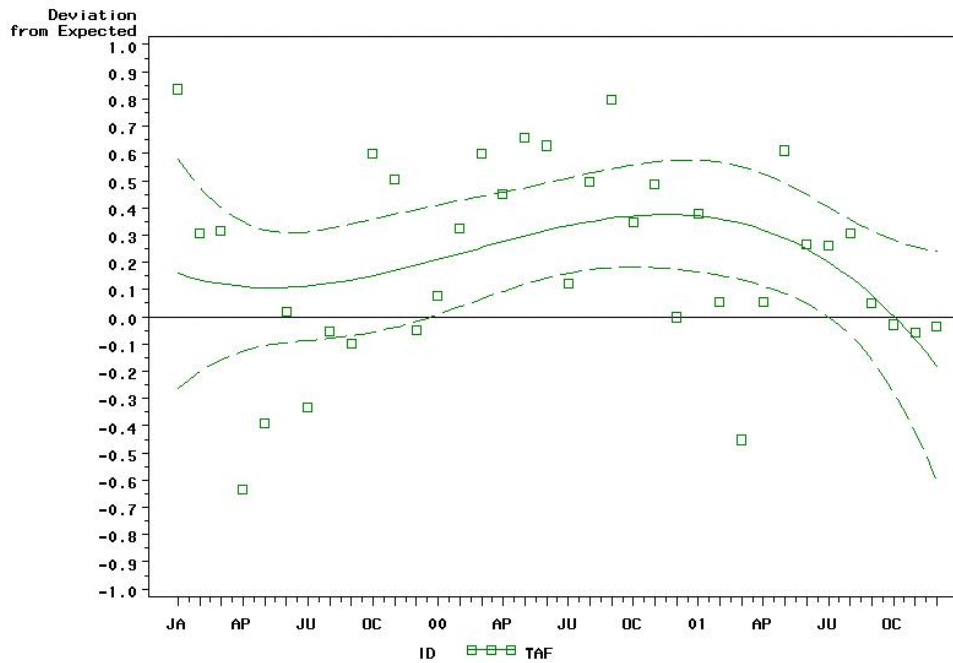


Figure 2.10

Other Misconduct at BOP Low—Security Prisons
Multilevel Models, January 1999 to December 2001

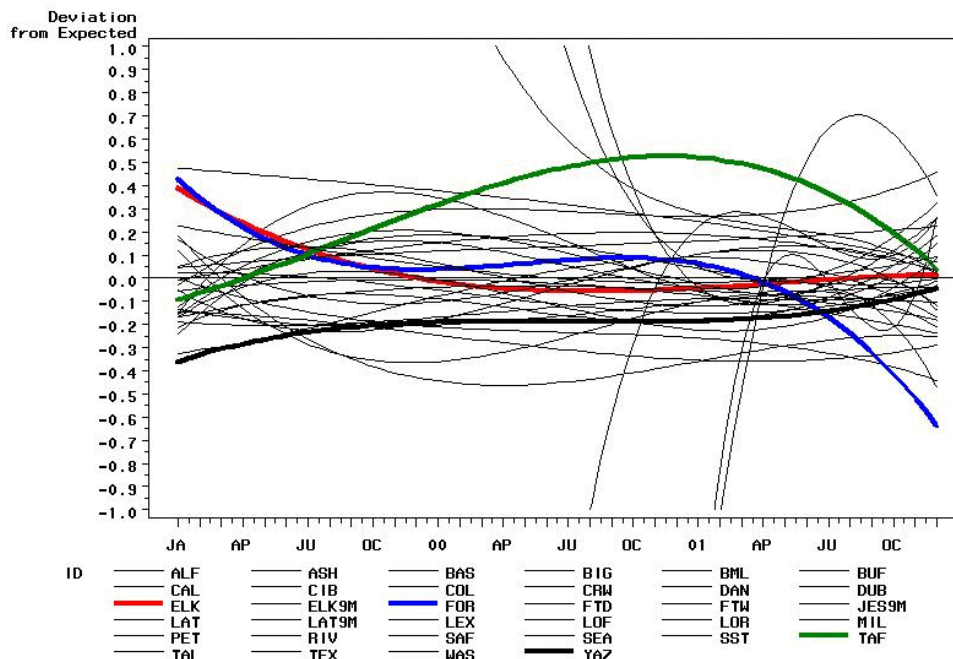


Figure 2.11

Other Misconduct at TCI and Comparison Prisons

Multilevel Models, January 1999 to December 2001

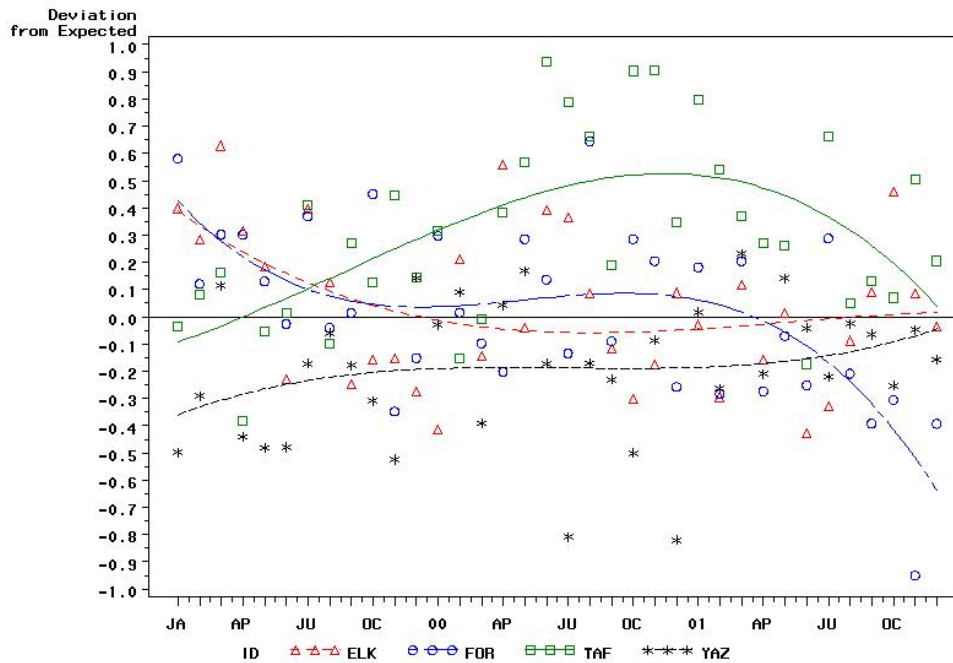
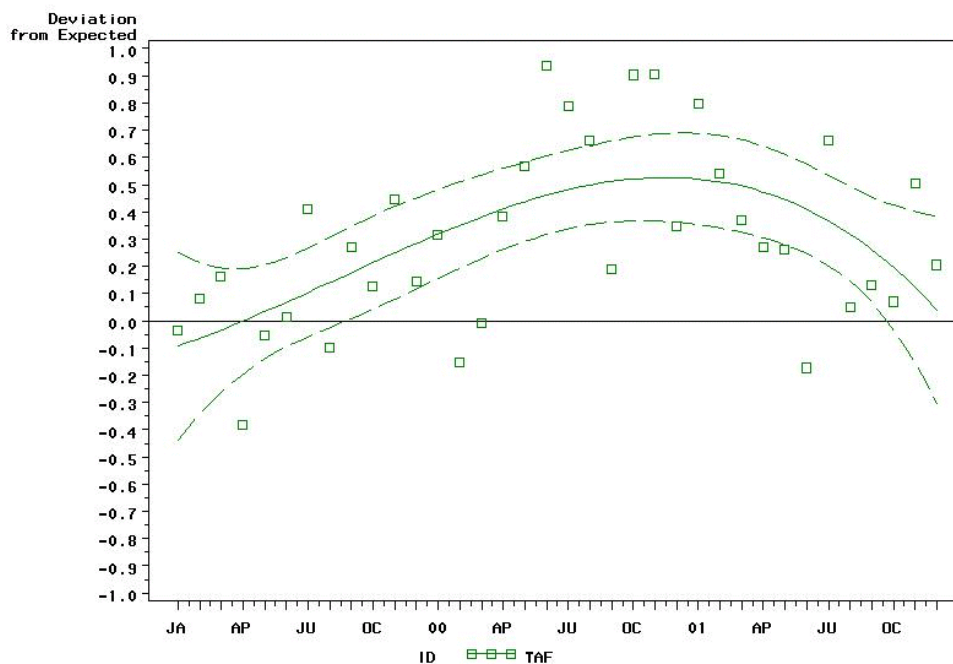


Figure 2.12

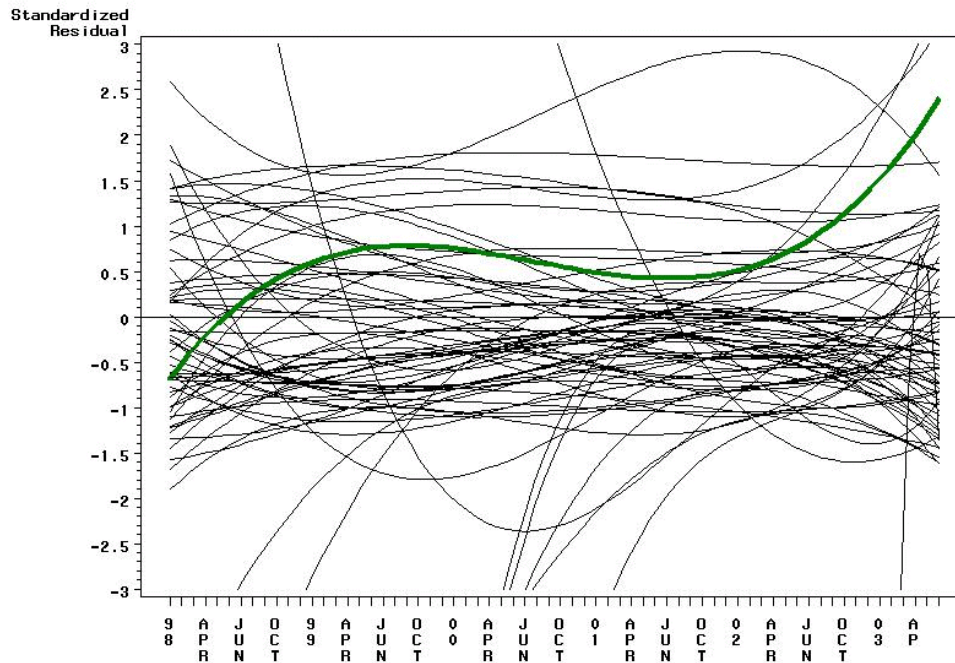
Other Misconduct at TCI, with 95% Confidence Interval

Multilevel Models, January 1999 to December 2001

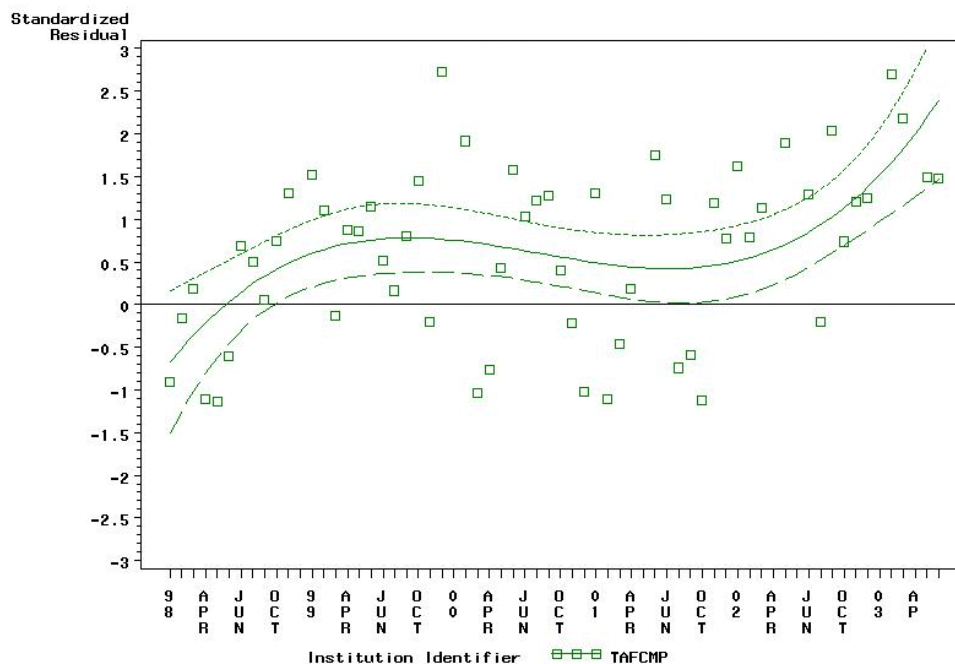


APPENDIX 3. MISCONDUCT PERFORMANCE MEASURES FOR TAFT CAMP

All Misconduct at BOP Minimum—Security Prisons



All Misconduct at TCI Camp, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003



100—Level Misconduct at BOP Minimum—Security Prisons
Aggregate Only Models, January 1998 to June 2003



Figure 3.7

300/400—Level Misconduct at BOP Minimum—Security Prisons
Aggregate Only Models, January 1998 to June 2003

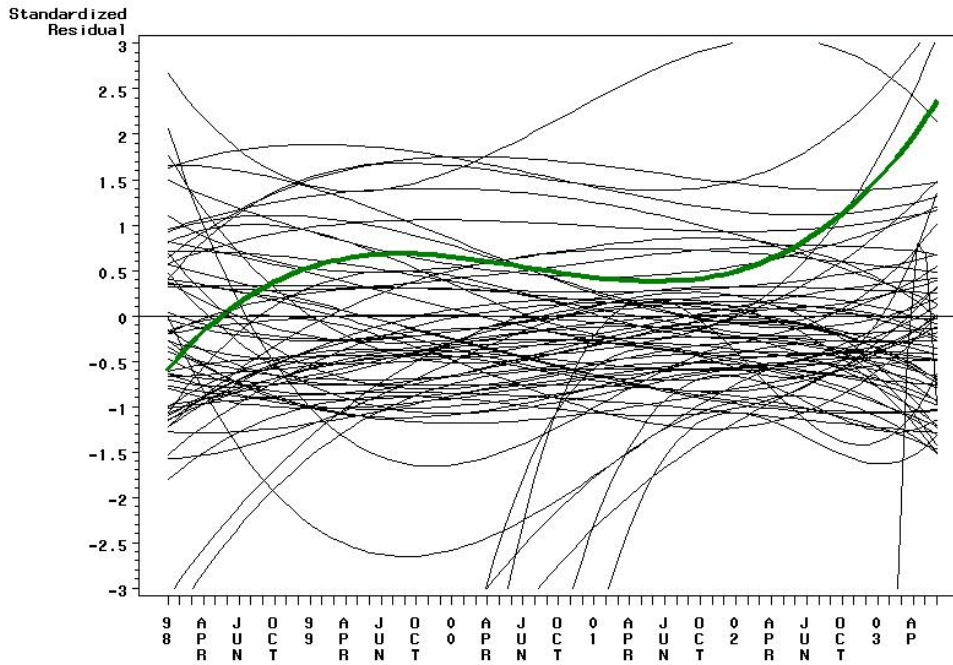
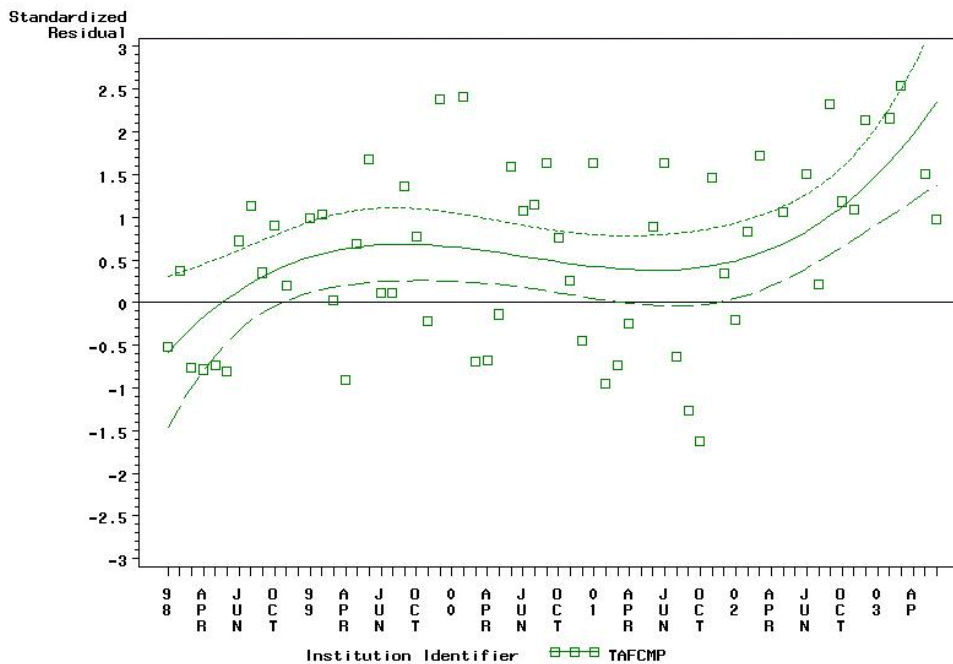


Figure 3.8

300/400—Level Misconduct at TCI Camp, with 95% Confidence Interval
Aggregate Only Models, January 1998 to June 2003



REFERENCES

- Andrews, R. D. (1999). *Food and Work Strike, August 25, 1999: After Action Review*. Taft, CA: Taft Correctional Institution.
- Archambeault, W. G., & Deis, D. R. (1996). *Cost Effectiveness Comparisons of Private Versus Public Prisons in Louisiana: A Comprehensive Analysis of Allen, Avoyelles, and Winn Correctional Centers*. Baton Rouge: Louisiana State University.
- Austin, J., & Coventry, G. (2003). A Second Look at the Private Prison Debate. *The Criminologist*, 28(5), 1, 3-9.
- Blumstein, J. F., & Cohen, M. A. (2003). *The Interrelationship Between Public and Private Prisons: Does the Existence of Prisoners Under Private Management Affect the Rate of Growth in Expenditures on Prisoners Under Public Management*. Alexandria, VA: Corrections Corporation of American and Association for Private Correctional and Treatment Organizations.
- Britton, D. M. (1997). Perceptions of the Work Environment Among Correctional Officers: Do Race and Sex Matter? *Criminology*, 35(1), 85.
- Camp, S. D. (1999). Do Inmate Survey Data Reflect Prison Conditions? Using Surveys to Assess Prison Conditions of Confinement. *Prison Journal*, 79(2), 250-268.
- Camp, S. D., & Gaes, G. G. (2000). *Private Prisons in the United States, 1999: An Assessment of Growth, Performance, Custody Standards, and Training Requirements*. Washington, D.C.: Federal Bureau of Prisons.
- Camp, S. D., & Gaes, G. G. (2001). Private Adult Prisons: What Do We Really Know and Why We Don't Know More? In D. Shichor & M. Gilbert (Eds.), *Privatization of Criminal Justice: Past, Present and Future* (pp. 283-298). Cincinnati, OH: Anderson.
- Camp, S. D., & Gaes, G. G. (2002). Growth and Quality of U.S. Private Prisons: Evidence from a National Survey. *Criminology and Public Policy*, 1(3), 427-449.
- Camp, S. D., Gaes, G. G., Klein-Saffran, J., Daggett, D. M., & Saylor, W. G. (2002). Using Inmate Survey Data in Assessing Prison Performance: A Case Study Comparing Private and Public Prisons. *Criminal Justice Review*, 27(1).
- Camp, S. D., Gaes, G. G., Langan, N. P., & Saylor, W. G. (2003). The Influence of Prisons on Inmate Misconduct: A Multilevel Investigation. *Justice Quarterly*, 20(3), 501-533.

- Camp, S. D., Gaes, G. G., & Saylor, W. G. (2002). Quality of Prison Operations in the Federal Sector: A Comparison with a Private Prison. *Punishment & Society*, 4(1), 27-53.
- Camp, S. D., Saylor, W. G., & Gaes, G. G. (2001). *Comparing Private and Public Prisons Using Operational Data: Correcting the Apples to Oranges Comparisons*. Washington, D.C.: Federal Bureau of Prisons.
- Camp, S. D., Saylor, W. G., & Harer, M. D. (1997). Aggregating Individual-Level Evaluations of the Organizational Social Climate: A Multilevel Investigation of the Work Environment at the Federal Bureau of Prisons. *Justice Quarterly*, 14(4), 739-761.
- Camp, S. D., Saylor, W. G., & Wright, K. N. (1999). Creating Performance Measures from Survey Data: A Practical Discussion. *Corrections Management Quarterly*, 3(1), 71-80.
- Comptroller of the Treasury. (2003). *Performance Audit: Department of Correction*. Nashville, TN: Comptroller of the Treasury, Division of State Audit.
- Crane, R. (2001). *Monitoring correctional services provided by private firms*. Washington, D.C.: Correctional Programs Office, U. S. Department of Justice.
- Crants, D. R., III. (1991). Private Prison Management: A Study in Economic Efficiency. *Journal of Contemporary Criminal Justice*, 7(1), 49-59.
- Gaes, G. G., Camp, S. D., Nelson, J., & Saylor, W. G. (2004). *Prison Performance: Laying the Groundwork to Compare Public and Private Prisons*. Lanham, MD: Alta Mira Press.
- Gaes, G. G., Camp, S. D., & Saylor, W. G. (1998). Appendix 2: Comparing the Quality of Publicly and Privately Operated Prisons: A Review. In D. McDonald, E. Fournier, M. Russell-Einhorn & S. Crawford (Eds.), *Private Prisons in the United States: An Assessment of Current Practice* (pp. 1-38). Boston, MA: Abt Associates Inc.
- General Accounting Office. (1996). *Private and Public Prisons: Studies Comparing Operational Costs and/or Quality of Service*. Washington, D.C.: U.S. General Accounting Office.
- Goldstein, H. (1995). *Multilevel Statistical Models* (Second ed.). New York: John Wiley.
- Harding, R. W. (1997). *Private Prisons and Public Accountability*. New Brunswick, NJ: Transaction.
- Harding, R. W. (2001). Private Prisons. In M. Tonry (Ed.), *Crime and Justice: A Review of Research* (Vol. 28, pp. 265-346). Chicago: University of Chicago Press.
- Harrison, P. M., & Beck, A. J. (2003). *Prisoners in 2002*. Washington, D.C.: Bureau of Justice Statistics.

- Hart, B. (2003). Private-prison Push Stalls in State. *The Arizona Republic*.
- Hartigan, J. A. (1975). *Clustering Algorithms*. New York: John Wiley & Sons.
- Heinrich, C. J., & Lynn, L. E. (2000). Means and Ends: A Comparative Study of Empirical Methods for Investigating Governance and Performance. *Journal of Public Administration and Theory*, 11(1), 109-138.
- Howell, S. (2001). Parallel Coordinates Display (Version 2). Washington, D.C.: Federal Bureau of Prisons.
- Kreft, I., & De Leeuw, J. (1998). *Introducing Multilevel Models*. Thousand Oaks: Sage Publications.
- Lanza-Kaduce, L., Parker, K., & Thomas, C. W. (2000). The Devil in the Details: The Case Against the Case Study of Private Prisons, Criminological Research, and Conflict of Interest. *Crime & Delinquency*, 46(1), 92-136.
- Lincoln, J. R., & Zeitz, G. (1980). Organizational Properties from Aggregate Data: Separating Individual and Structural Effects. *American Sociological Review*, 45(3), 391-408.
- Logan, C. H. (1990). *Private Prisons: Cons and Pros*. New York: Oxford University Press.
- Logan, C. H. (1992). Well Kept: Comparing Quality of Confinement in Private and Public Prisons. *The Journal of Criminal Law and Criminology*, 83(3), 577-613.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.
- McDonald, D. C., Fournier, E., Russell-Einhorn, M., & Crawford, S. (1998). *Private Prisons in the United States: An Assessment of Current Practice*. Boston, MA: Abt Associates Inc.
- McDonald, D. C., & Patten, C., Jr. (2003). *Governments' Management of Private Prisons*. Cambridge, MA: Abt Associates Inc.
- McDonald, D. C., Patten, C., Jr., Heliotis, J., Zeytoonjian, L., Kling, R., & Carlson, K. (2002). *A Study to Compare Costs and Performance at Taft Correctional Institution and Government-Operated Federal Prisons*. Cambridge, MA: Abt Associates Inc.
- Moore, A. T. (1998). *Private Prisons: Quality Corrections at Lower Cost*. Los Angeles: Reason Public Policy Institute.
- Nathan, S. (2000, July 2000). North Carolina Leads the Way. *Prison Privatization Report International*, 36, 1-2.
- Nelson, J. (1998). Appendix 1: Comparing Public and Private Prison Costs. In D. McDonald, E. Fournier, M. Russell-Einhorn & S. Crawford (Eds.), *Private Prisons in the United States: An Assessment of Current Practice* (pp. 1-34). Boston, MA: Abt Associates Inc.

- Nelson, J. (1999). *Taft Prison Facility: Cost Scenarios*. Washington, DC: National Institute of Corrections.
- Nelson, J. (2005). *Competition in Corrections: Comparing Public and Private Sector Operations*. Alexandria, VA: The CNA Corporation.
- Normand, S.-L., Glickman, M. E., & Gatsonis, C. A. (1997). Statistical Models for Profiling Providers of Medical Care: Issues and Application. *Journal of the American Statistical Association*, 92(439), 803-814.
- Perrone, D., & Pratt, T. C. (2003). Comparing the Quality of Confinement and Cost-Effectiveness of Public versus Private Prisons: What We Know, Why We Do Not Know More, and Where to Go from Here. *The Prison Journal*, 83(3), 301-322.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second Edition ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2001). HLM for Windows 5.0: Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International.
- SAS International. (2002). SAS 8.02. Cary, NC: SAS International.
- Saylor, W. G. (1984). *Surveying Prison Environments*. Washington, D.C.: Federal Bureau of Prisons.
- Saylor, W. G. (1988). *Developing a Strategic Support System: Monitoring the Bureau's Performance via Trends in Key Indicators*. Washington, D.C.: Federal Bureau of Prisons.
- Saylor, W. G. (1989). Quality Control for Prison Managers. *Federal Prisons Journal*, Fall, 39-42.
- Saylor, W. G. (1992). *The Role of the Key Indicators/Strategic Support System in the Relationship between Quality Control Assurance and Strategic Planning*. Washington, D.C.: Federal Bureau of Prisons.
- Saylor, W. G., & Wright, K. N. (1992). Status, Longevity, and Perceptions of the Work Environment among Federal Prison Employees. *Journal of Offender Rehabilitation*, 17(3/4), 133-160.
- Segal, G. F., & Moore, A. T. (2002). *Weighing the Watchmen: Evaluating the Costs and Benefits of Outsourcing Correctional Services*. Los Angeles: Reason Public Policy Institute.
- Shichor, D. (1995). *Punishment for Profit: Private Prisons/Public Concerns*. Thousand Oaks, CA: Sage.

- Thomas, R. C., Gookin, K., Keating, B., Whitener, V., Williams, R. M., Crane, R., et al.
(1996). *Department of Corrections Privatization Feasibility Study* (No. 96-2):
Legislative Budget Committee, State of Washington.
- Wegman, E. (1990). Hyperdimensional Data Analysis using Parallel Coordinates. *Journal of
the American Statistical Association*, 85(411), 664-675.
- Zelterman, D. (2002). *Advanced Log-Linear Models using SAS*. Cary, NC: SAS Institute, Inc.